NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

AN ANALYSIS OF SECURITY BACKGROUND
BACKGROUND INVESTIGATION DATA WITH
RELATION TO SUBSEQUENT DISCHARGE

By

Edward R. Koucheravy

September 1988

Thesis Advisor: Peter A. W. Lewis

Approved for public release; distribution is unlimited

:lassified

ity classification of this page

REPORT DOCUMENTATION PAGE

Report Security Classification Unclassified		1b Restrictive Markings	
Security Classification Authority		3 Distribution Availability of Report	
Declassification Downgrading Schedule		Approved for public release; distribution is unlimited.	
Performing Organization Report Number(s)		5 Monitoring Organization Report Number(s)	
Name of Performing Organization val Postgraduate School	6b Office Symbol *(if applicable)* 55	7a Name of Monitoring Organization Naval Postgraduate School	
Address *(city, state, and ZIP code)* nterey, CA 93943-5000		7b Address *(city, state, and ZIP code)* Monterey, CA 93943-5000	
Name of Funding Sponsoring Organization	8b Office Symbol *(if applicable)*	9 Procurement Instrument Identification Number	
Address *(city, state, and ZIP code)*		10 Source of Funding Numbers	
		Program Element No / Project No / Task No / Work Unit Accession No	

Title *(Include security classification)* AN ANALYSIS OF SECURITY BACKGROUND INVESTIGATION DATA AND E RELATIONSHIP WITH SUBSEQUENT DISCHARGE

Personal Author(s) Edward R. Koucheravy

Type of Report	13b Time Covered	14 Date of Report *(year, month, day)*	15 Page Count
ster's Thesis	From To	September 1988	116

Supplementary Notation The views expressed in this thesis are those of the author and do not reflect the official policy or po-on of the Department of Defense or the U.S. Government.

Cosati Codes			18 Subject Terms *(continue on reverse if necessary and identify by block number)*
l	Group	Subgroup	Cross-Tabulation,Chi-Square Test, Security Investigation,

Abstract *(continue on reverse if necessary and identify by block number)*

This thesis is concerned with the analysis of security investigation data extracted from the investigative files of 564 U.S. y first-term enlisted personnel who came on active duty between 1979 and 1982. The individuals had all completed their term of service and had either completed service satisfactorily or had been released early with an adverse discharge. The a was selected from six character-of-service categories: good, homosexual, drug/alcohol abuse, misconduct, court martial, character and behavior disorders. The purpose of the thesis was to investigate optimal ways to configure a large, cate-cal data base and to look for and quantify relationships between investigative data and final disposition of service. Several eworthy relationships were found between derogatory information developed in the investigation and the subsequent racter-of-service. Further avenues of investigation using this data are suggested.

Distribution Availability of Abstract nclassified unlimited ☐ same as report ☐ DTIC users	21 Abstract Security Classification Unclassified	
Name of Responsible Individual r A. W. Lewis	22b Telephone *(include Area code)* (408) 646-2283	22c Office Symbol 55Lw

FORM 1473,84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete

security classification of this page

Unclassified

Approved for public release; distribution is unlimited.

An Analysis of Security Background Investigation Data and the Relationship With Subsequent Discharge

by

Edward R. Koucheravy
Captain, United States Army
B.S., United States Military Academy, 1978

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1988

ABSTRACT

This thesis is concerned with the analysis of security investigation data extracted from the investigative files of 564 U.S. Navy first-term enlisted personnel who came on active duty between 1979 and 1982. The individuals had all completed their first term of service and had either completed service satisfactorily or had been released early with an adverse discharge. The data was selected from six character-of-service categories: good, homosexual, drug/alcohol abuse, misconduct, court martial, and character and behavior disorders. The purpose of the thesis was to investigate optimal ways to configure a large, categorical data base and to look for and quantify relationships between investigative data and final disposition of service. Several noteworthy relationships were found between derogatory information developed in the investigation and the subsequent character-of-service. Further avenues of investigation using this data are suggested.

TABLE OF CONTENTS

I. INTRODUCTION

A. BACKGROUND

The importance of protecting sensitive military information and operations from potentially hostile sources is a concept as old as warfare itself. Events of the recent past indicate that the nation must never grow complacent about its ability to safeguard classified information. World-wide defense commitments, the ideological and historic differences existing between the US and other nations, and the huge number of people who frequently access, create, analyze and service the vast amount of sensitive information combine to create a tremendous managerial problem: Who can be trusted with access to the nation's security secrets?

The need to investigate the backgrounds of those people needing access to classified information has been a fixture of the national security establishment for many years. Typically, an individual, by virtue of his duty responsibilities, is determined to need regular access to sensitive information of some level (secret, top secret, sensitive compartmentalized information, etc). A fairly standard administrative procedure is employed throughout the Department of Defense (DOD) in order to determine whether the person should be allowed access to classified information,

B. THE SECURITY INVESTIGATION PROCEDURE

The first element in a security investigation is the completion of a detailed form named the Statement of Personal History (SPH). The SPH requires specific information about a person's past. Information such as a list of close family members, foreign travel, arrests and convictions, schools attended, jobs held, creditors, and personal references are all required. The SPH is the starting point for any security investigation.

The next step in the investigation consists of the National Agency Check (NAC) and the Local Agency Check (LAC). Law enforcement agencies, both local (i.e., city or state police) and national (i.e., the FBI) are queried about outstanding warrants and records of arrests. A check of credit information is also conducted with national and local credit bureaus to determine whether an individual has money problems.

The clearance will normally be granted to a person who requires access to information with a classification of Secret or lower when the above procedure does not turn up any inconsistencies.

A person requiring access to top secret or higher level information will undergo a much more detailed investigation: a background investigation (BI), or a special background investigation (SBI). These investigations are much more thorough than those for lesser clearances and involve actual interviews with people who know and have developed a relationship with the individual being investigated. Neighbors, friends, school officials, former employers and others may be interviewed. If the answers are consistent and positive, the subsequent investigation will be much less detailed than if a negative trend develops and other sources of information are "developed" by the investigators. If information is developed which contradicts that listed on the Statement of Personal History or is conspicuously absent from it, the subject will almost certainly be interviewed. In certain other types of investigations, an interview is always required.

The result of this investigation is a dossier containing basic biographical data, derogatory information obtained from the SPH and other sources (or lack of such information) and recommendations as to the trustworthiness of the subject of the investigation. Derogatory information varies from traffic infractions to emotional problems to felonies. All the investigative data is gathered for the clearance determination. An adjudictor reads the investigation file and makes the judgement as to the award of the clearance.

The last step in the security investigation process is a review of the information obtained and determination of whether the clearance should be granted.

Review of the information is performed in accordance with Adjudication guidelines contained in the DOD Personal Security Regulation, DOD 5200.2-R, dated January, 1987. The factors which can disqualify an individual for a clearance are listed as well as the mitigating factors which might allow a clearance to be granted even though a disqualifying factors are present in the information. For example, a person might admit to experimental use of marijuana (less than six instances of use) in their adolescence. This use of cannabis (marijuana or its derivatives) is considered a disqualifying factor. A mitigating factor in this instance is that the experimental abuse occurred more than six months ago, and the individual has no intention of using cannabis or other drugs in the future [Ref. 1].

The final determination of clearance for an individual whose record contains disqualifying information is a subjective one. It is based upon the merits of the case, and the evaluation of the adjudicator as to the mitigating factors which hopefully indicate the actual reliability of the individual in the future.

C. BACKGROUND OF THE SPECIAL BACKGROUND INVESTIGATION DATA BASE (SBID)

It is apparent that the investigation procedure must generate a tremendous amount of data about every person who is investigated for a security clearance. It is clear that we do not wish to trust national security information to those who are untrustworthy enough to violate laws, regulations, and accepted standards of conduct. Could this data be used to examine whether data obtained from the security investigations were in any way related to the future service record of those investigated? Could this data provide insight into the investigation process, allowing investigative resources to be more efficiently allocated?

The Defense Personal Security Research and Education Center (PERSEREC) in Monterey, California was directed to examine a large sample of data produced from security investigations of first-term enlistees entering the Navy during the years 1979 - 1982. The purpose of the study was to develop insight about the information developed in security investigations, especially when the final disposition of service of investigative subjects was known.

The individuals whose records were involved in the study:

1. Had background investigations initiated within three months of enlistment;
2. Were separated or discharged during, or upon completion of their initial tour of duty;
3. Were discharged for homosexuality, misconduct, drug abuse, court martial, character and behavior disorder, or normal completion of enlistment.

Thus, in the data base, there are five types of unsuitability discharge categories and one control group of personnel who successfully completed their term of service.

Seven-hundred records were selected randomly (based upon the last digit of the social security number) for the study. One-hundred cases were selected from each of the five unsuitability discharge groups and two-hundred cases in which the individuals were normally separated. The number of cases which were eventually included in the study numbered 564 because those cases where the Background Investigation was cancelled for any reason were removed.

The number of records chosen in each category were not in relation to the character-of-service category's proportion in the actual population. An immense number of records would need to be drawn as a single sample in order to get a large enough representation from each adverse discharge category. As an illustration, consider that there

are 73 records in this data base from the court martial character of service category. Persons who are investigated receive this adverse character-of-service designation approximately 0.18% of the time. Simple arithmetic indicates that to get approximately 73 records in this category from a single sample from the investigation population at large would require a sample size of nearly 41,000. It seems obvious that this is not reasonable. Table 1 displays the approximate percentages of those initially investigated who receive each of the six character-of-service designations discussed in this thesis [Ref. 2]. There are other designations which are not considered here.

Table 1. CHARACTER OF SERVICE CATEGORY PROPORTIONS

Character of Service Category	% of Investigation Population Receiving Category
Good	90.4%
Homosexual	0.92%
Misconduct	1.2%
Drug/Alcohol Abuse	1.8%
Court Martial	0.18%
Character Behavior Disorder	0.65%

The data base was created by taking the investigation information from microfiche and entering it into a Lotus 123 spreadsheet. There were 93 possible entries for each of the 564 records resulting in a total data base with the potential for approximately 52,500 data points.

The data was essentially categorical in nature with an individual record containing personal information ranging from date of birth and military specialty to findings from high school to type of discharge. A four-digit code representing the type of derogatory information was the prime means of listing this data and allowed standardization across the data base. Other codes were created to represent other pieces of information such as the recommendations obtained at the various sources (high schools, colleges, neighborhoods, etc.), race or marital status.

Problems with the size of the data base, the slow response of an AT-style microcomputer when dealing with such a large data set, and the limitations of Lotus 123 in performing statistical functions allowed only a cursory analysis of the data base as

originally implemented. Clearly another approach was necessary to analyze and obtain insights from this data.

D. PURPOSE

The purpose of this thesis is two-fold: to investigate some available methods for organizing and analyzing a large, categorical data base; and to use statistical and data-analytical techniques to evaluate the personal security data detailed above in order to develop insights and correlations between the security investigation data and the subsequent disposition of the subject's term of enlistment.

E. LIMITATIONS

The data used in this paper was analyzed as provided. It was not possible to ensure actual random selection of the data, however we assume that each sample was selected randomly. The data was selected in an arbitrary manner (one-hundred records from each of the unsuitability discharge categories and two-hundred records with normal completion of service). It may be difficult to apply the results of this investigation to the general population.

F. ANALYTICAL TOOLS USED

The data was initially reduced and documented using the Statgraphics (version 2.6) statistical software package on a Compaq 286 portable personal computer with two megabytes of additional random access memory (RAM). After reduction it was transferred to an IBM 3033 System 370 mainframe computer using the MVS batch system. On the mainframe computer, Grafstat, an unreleased IBM mainframe data analysis and statistical package was used. In addition, APL programs for categorical analysis were written using APL Graphpak to supplement the routines available in Grafstat.

G. ORGANIZATION OF THESIS

Following this introduction, the data reduction techniques used for this thesis and the lessons learned from that effort are discussed in Chapter II. The main body of the thesis is contained in the Chapter III and deals with the data operations and the analysis conducted. Chapter IV discusses some promising areas for further analysis which were only briefly pursued because of time constraints. The closing chapter will summarize the results of this research, set forth the conclusions drawn from those results and provide recommendations for future research involving this data.

II. DATA REDUCTION

A. GENERAL

PERSEREC experienced problems in attempting to analyze a data base of this magnitude. This led them to investigate other methods of configuring the data in order to perform the analysis they felt was necessary. Subsequently, the Lotus 123 files were exported to the mainframe computer and configured into Conversational Monitoring System (CMS) ASCII files. The categorical nature of the data and its overwhelming size dictated that documentation and verification of the data base was necessary before any further useful analysis could be performed. However, the data editors available in CMS on the mainframe computer did not offer the ability to easily operate on column fields and did not have the flexibility needed to simultaneously document the work performed as it proceeded.

B. DATA EDITING

Statgraphics (version 2.6) offered a user-friendly data editor offering the requisite capabilities. Unfortunately, it was available only on a personal computer. A Compaq 286 portable AT-compatible micro-computer with two megabytes of additional memory (useable as a virtual disk) was used. It proved extremely useful; however, its size limited the amount of data which could be operated upon without exceeding the memory limitations of the computer (these memory restrictions will be alleviated in the future when using the new 80386 based machines).

The CMS files were transferred into micro-computer ASCII files and then stored on floppy disks and subsequently read into six Statgraphics (ASF) files. Each of the files consisted of approximately 15 of the variable entries for each of the 564 records (approximately 8400 data points). At any one time six or seven of these variables could be operated upon within the data editor.

A general procedure was followed in formatting and verifying each of the six files. First, the file was checked to insure that the data, as it existed on the CMS files, had been transferred correctly. In one instance half of the field of one variable was truncated and had to be reconstructed.

Next, the numeric coding used for each column was researched and ambiguities resolved by recoding or removal. This step required considerable research into the coding

methods and the investigation process in order to understand, and, if necessary, change the numeric codes for the sake of clarity.

Finally, a frequency tabulation of each column was performed and labels were created which corresponded to the coded values. These labels were especially useful later in the analysis when cross-tabulations between variables vectors were conducted.

The procedure discussed above was iterative as sometimes several interpretations resulted before one was confirmed as correct. Documentation of the data base was conducted throughout these three steps. The list of the variables contained in the data base, their purpose and their types are contained in Figure 1 through Figure 3 . These figures are a direct copy of the file management screen that appears in Statgraphics as you enter the full-screen editor or view the data directory. Comments are limited to 21 characters for each variable.

VARIABLE	WIDTH	TYPE	RANK	LENGTH	DATE	TIME	COMMENT
A	5	I	1	564	3/18/88	11:59	RECORD NO. (RANDOM)
C	3	I	1	564	2/26/88	11:08	SEX (MALE OR FEMALE)
D	8	D	1	564	3/18/88	13:02	BIRTHDATE
F	8	D	1	564	3/18/88	14:01	DATE OF ENTNAC
G	8	D	1	564	3/18/88	14:01	BI REQUEST DATE
I	3	I	1	564	2/26/88	11:08	REASON FOR BI
J	4	I	1	564	2/26/88	11:08	OCCUPATION CODE
K	3	I	1	564	2/26/88	11:10	REASON FOR INTERVIEW
L1	6	I	1	564	2/26/88	12:29	INTERVIEW INFO - 1.
L2	6	I	1	564	2/26/88	12:29	INTERVIEW INFO - 2.
L3	6	I	1	564	2/26/88	12:29	INTERVIEW INFO - 3.
L4	6	I	1	564	2/26/88	12:29	INTERVIEW INFO - 4.
M1	6	I	1	564	2/26/88	12:29	FBI/DCII FINDINGS1
M2	6	I	1	564	2/26/88	12:29	FBI/DCII FINDINGS2
N1	6	I	1	564	2/26/88	14:03	LOCAL AGENCY CHECK
N2	6	I	1	564	2/26/88	14:31	LOCAL AGENCY CHECK
N3	6	I	1	564	2/26/88	14:03	LOCAL AGENCY CHECK
N4	6	I	1	564	2/26/88	14:03	LOCAL AGENCY CHECK
O1	6	I	1	564	2/26/88	14:03	CREDIT BUREAU CHECK
O2	6	I	1	564	2/26/88	14:03	CREDIT BUREAU CHECK
P	4	I	1	564	2/26/88	10:59	H S - # OF SOURCES

Figure 1. List of Variables Contained in the Data Base: Extracted from the Statgraphics Data Management Screen.

VARIABLE	WIDTH	TYPE	RANK	LENGTH	DATE	TIME	COMMENT
Q1	3	I	1	564	3/ 4/88	14:32	HIGH SCHOOL RECOMM.
Q2	6	I	1	564	3/ 4/88	14:32	HIGH SCHOOL RECOMM.
Q3	6	I	1	564	3/ 4/88	14:32	HIGH SCHOOL RECOMM.
R1	3	I	1	564	2/26/88	15:47	HIGH SCHOOL FINDINGS
R2	5	I	1	564	2/26/88	16:30	HIGH SCHOOL FINDINGS
R3	5	I	1	564	2/26/88	16:30	HIGH SCHOOL FINDINGS
R4	5	I	1	564	2/26/88	16:30	HIGH SCHOOL FINDINGS
S	3	I	1	564	2/26/88	10:59	COLL. - # OF SOURCES
T	4	I	1	564	2/26/88	11:00	COLL. RECOMMENDATION
U	5	I	1	564	2/26/88	11:00	COLLEGE FINDINGS
V	3	I	1	564	2/26/88	11:00	EMPL. # OF SOURCES
W	3	I	1	564	2/26/88	10:53	CO-WORKER # SOURCES
X1	3	I	1	564	3/ 4/88	11:28	EMPLOYMENT RECOMM.
X2	6	I	1	564	3/ 4/88	11:15	EMPLOYMENT RECOMM.
X3	6	I	1	564	3/ 4/88	11:15	EMPLOYMENT RECOMM.
Y1	5	I	1	564	3/ 4/88	13:36	EMPLOYMENT FINDINGS
Y2	5	I	1	564	3/ 4/88	13:36	EMPLOYMENT FINDINGS
Y3	5	I	1	564	3/ 4/88	13:36	EMPLOYMENT FINDINGS
Y4	5	I	1	564	3/ 4/88	13:36	EMPLOYMENT FINDINGS
Z	2	I	1	564	2/26/88	10:54	NEIGH. # OF SOURCES
AA1	3	I	1	564	3/ 4/88	12:02	SPH NEIGH. RECOMM.
AA2	3	I	1	564	3/ 4/88	12:02	DEV. NEIGH. REC.
AA3	6	I	1	564	3/ 4/88	12:02	DEV. NEIGH. REC.
AB1	5	I	1	564	3/ 4/88	14:06	NEIGH. FINDINGS
AB2	5	I	1	564	3/ 4/88	14:06	NEIGH. FINDINGS
AB3	5	I	1	564	3/ 4/88	14:06	NEIGH. FINDINGS
AC	3	I	1	564	2/26/88	10:56	# OF OTHER SOURCES
AD1	3	I	1	564	3/ 4/88	15:17	OTHER RECOMM.
AD2	6	I	1	564	3/ 4/88	15:17	OTHER RECOMM.
AE1	5	I	1	564	3/11/88	08:29	OTHER FINDINGS
AE2	5	I	1	564	3/11/88	08:29	OTHER FINDINGS
AE3	5	I	1	564	3/11/88	08:29	OTHER FINDINGS
AE4	5	I	1	564	3/11/88	08:29	OTHER FINDINGS
AF	2	I	1	564	2/26/88	10:46	RACE

Figure 2. List of Variables Contained in the Data Base (Continued): Extracted from the Statgraphics Data Management Screen.

VARIABLE	WIDTH	TYPE	RANK	LENGTH	DATE	TIME	COMMENT
AG	2	I	1	564	2/26/88	10:48	MARITAL STATUS
AJ	2	I	1	564	2/26/88	10:48	DEPENDENTS
AN	6	I	1	564	2/26/88	10:48	# OF SIBLINGS
AO	3	I	1	564	2/26/88	10:48	PERMANENT RESIDENCE
AQ	7	I	1	564	3/18/88	14:17	ENLISTMENT DATE
AR	5	1	1	564	2/26/88	10:48	AGE AT ENLISTMENT
AS	4	I	1	564	2/26/88	10:49	MONTHS HS TO ENLIST
AT	4	I	1	564	2/26/88	10:49	# JOBS HS TO ENLIST
AU	3	I	1	564	2/26/88	10:49	# MONTHS UNEMPL.
AV	3	I	1	564	2/26/88	10:49	# MONTHS COLLEGE
AW	3	I	1	564	2/26/88	10:49	MO.UNEMPL. PRIOR ENL
AX1	5	I	1	564	3/11/88	12:54	UNFAV. INFO. ON SPH
AX2	5	I	1	564	3/11/88	12:54	UNFAV. INFO. ON SPH
AX3	5	I	1	564	3/11/88	12:54	UNFAV. INFO. ON SPH
AX4	5	I	1	564	3/11/88	12:54	UNFAV. INFO. ON SPH
AY1	5	I	1	564	3/11/88	11:36	SUMMARY BI
AY2	5	I	1	564	3/11/88	11:36	SUMMARY BI
AY3	5	I	1	564	3/11/88	11:36	SUMMARY BI
AY4	5	I	1	564	3/11/88	11:36	SUMMARY BI
BB		C	2	564 8	3/18/88	14:42	
BC	3	I	1	564	2/26/88	10:33	CLEARANCE TYPE
BD		C	2	564 8	3/18/88	14:43	CLEARANCE REV.: DATE
BE		C	2	564 8	3/18/88	14:44	DATE OF SEPERATION
BF	3	I	1	564	2/26/88	10:36	RELEASE CODE
BG1	5	I	1	564	3/11/88	14:06	MILITARY OFFENSES
BG2	5	I	1	564	3/11/88	14:06	MILITARY OFFENSES
BG3	5	I	1	564	3/11/88	14:06	MILITARY OFFENSES
BG4	5	I	1	564	3/11/88	14:06	MILITARY OFFENSES
BH1	5	I	1	564	3/11/88	14:29	REMARKS/DISCHARGE
BH2	5	I	1	564	3/11/88	14:30	REMARKS/DISCHARGE
BH3	5	I	1	564	3/11/88	14:30	REMARKS/DISCHARGE
BH4	5	I	1	564	3/11/88	14:30	REMARKS/DISCHARGE
BL	4	I	1	564	2/26/88	10:08	STATUS OF 5520/20
BM	5	I	1	564	2/26/88	10:08	DISCHARGE CASE CODE
BO	3	I	1	564	2/26/88	10:08	INTERSVC. SEP. CODE
BP	2	I	1	564	2/26/88	10:08	CHARACTER OF SERVICE
BQ	2	I	1	564	2/26/88	10:08	TYPE OF DISCHARGE

Figure 3. List of Variables Contained in the Data Base (Continued): Extracted from the Statgraphics Data Management Screen.

C. DATA REPRESENTATION PROBLEMS

Inherent in the verification and documentation of a large data base obtained from an outside source are coding inconsistencies. Ideally, thorough documentation of the

codes used and the thought process employed in creating the data base is included with it. However, this is seldom the case.

The PERSEREC data base had many inconsistencies along with several strengths. A major strength of the data organization was the standardization of most of the coding employed. Derogatory information codes (used in 43 of the 93 columns) and recommendation codes (used in 13 of the columns) were used in a fairly standard manner. The numeric code for all derogatory information contained in the data base consisted of a standard four-digit code representing 135 different infractions. The list of infractions and their codes is listed in Appendix B.

The numeric code used for the types of recommendations obtained from various sources consisted of a two-digit integer representing the total number of persons who:

1. Recommended the subject for a position of trust;
2. Recommended the subject for a position of trust, with supervision;
3. Did not recommend the subject for a position of trust;
4. Declined comment.

Most sources of derogatory information are represented by several columns in the data base. A source is considered a location such as college, high school, employer, neighborhood, etc.. Multiple columns are available for each source category to allow room for several different types of derogatory information to be displayed, if necessary. Table 2 shows how the information of columns Y1, Y2, Y3, and Y4 (findings or derogatory information obtained from employers) was represented:

Table 2. INITIAL REPRESENTATION OF DEROGATORY INFORMATION (EXAMPLE).

Record Number	Y1	Y2	Y3	Y4
1	9999	9999		
2	9999	1071	1106	
3	1829	9999	1844	9999
4	1805	1824		

After research, these records were interpreted in the following manner: If there are only 9999 entries in a particular record's entries in Y1 - Y4, then no derogatory

information from the subject's former employers was found. The possibility of no interview being conducted is reasonable, although all information indicates that former employers were visited in almost all instances. If any 9999 entries are contained along with derogatory information for a particular record, those 9999 codes are meaningless. The corrected records are shown in Table 3 .

Table 3. REPRESENTATION OF DEROGATORY INFORMATION AFTER REDUCTION (EXAMPLE).

Record Number	Y1	Y2	Y3	Y4
1	9999			
2	1071	1106		
3	1829	1844		
4	1805	1824		

In this table no information was obtained on the person represented by record number 1. For the second person, the investigator found evidence that the person was known to lie (1071), and that he was at some time intoxicated in public (1106). The third person had evidence of vandalism (1829) and malicious mischief (1844). The fourth person was found to have an incident of reckless driving (1805) and also illegal use of a firearm (1824).

Columns representing derogatory information obtained from colleges, high schools, neighbors, and other sources were similarly reduced.

As discussed above, the 9999 code used in columns Y1 - Y4 represented "no derogatory information." Research revealed that this interpretation of the 9999 code could not be used in some of the other columns. In the security investigation realm, employers and neighbors are considered "productive" sources. With that designation, the former employers and neighbors of a subject are almost always interviewed, thus the 9999 code for those sources means "no derogatory information." Sources other than employers and neighbors, on the other hand, are normally only visited by an investigator when he is fairly certain to obtain derogatory information. The 9999 code in conjunction with these types of sources means "no interview conducted."

An even more confusing coding scheme was discovered relating to the recommendations obtained from the five types of sources outlined above. For the employer, high

school, college and other sources, a 99 code represents "no interview." The coding for neighborhood recommendations was different.

Neighborhoods are the source of many developed sources of derogatory information. A distinction was made between the recommendations of neighbors listed on the SPH (generally positive) and those from neighborhood sources developed by the investigators. This resulted in four possible entries for recommendations from a subject's neighborhood. The column vectors representing information obtained from the subject's neighbor are designated AA1-AA4. Column AA1 represents the recommendations obtained from persons listed on a subject's Statement of Personal History. Entries in columns AA2-AA4 were recommendations obtained from neighborhood sources developed by the investigator. A 99 entry in column AA1 meant "no interview conducted," while a 99 entry in column AA2 means "no sources developed." Furthermore, a 99 entry in columns AA3 or AA4 meant nothing. These variable fields were repaired by removing all 99 codes from columns AA3 and AA4.

Another instance of miscoding occurred in column AN, which represents the number of siblings of the subject. Throughout the field a character code of "U" existed along with the usual integers (1, 2, ...) representing the number of siblings. This code was thoroughly researched until the only possible explanation was obtained--it represented "unknown."

The problems highlighted here point to the importance of differentiating, by coding, even small differences in meaning when implementing codes. The failure to do so risks losing important distinctions which may in fact invalidate the data. Another point to be made is that documentation is essential when data bases are created. Luckily, the person who performed the data entry was available for reference throughout the data reduction stage of this project, otherwise much of the information contained in the data base might have been lost.

Erroneous entries were not commonly found in the data base. Only two erroneous codes (not of the 135 actual derogatory information codes) were found and they were in the same column. Research into the underlying record revealed that the codes had digits transposed and the corrections were easily made.

Missing values, or blanks, were common in some columns. Care had to be taken to preserve these blanks when transferring from one system to another. The Statgraphics representation of blanks as the integer -32768 proved useful in this regard.

The files were initially represented in a random order by record number. This proved inconvenient when cross-validation of the record to its original file was necessary. The use of APL in conjunction with Statgraphics allowed all records to be reorganized in ascending order and made the file much easier to reference.

Date fields were entered as six-digit codes representing month-day-year. Problems were encountered with formatting as Statgraphics requires a slash (/) between the month and day and the day and year. A simple APL function was written which performed this conversion.

D. RECOMMENDATIONS FOR CODING A LARGE DATA BASE

1. Care must be taken to differentiate even subtle variations in meaning by using different codes.
2. The data base must be designed with the proper analytical tool (software and hardware) consistent with the purpose and goals of the analysis.
3. Proper documentation is essential when creating a data base. This is important not only for the data base creators to have for their own memory, but also so that others may use the data base. It is also important because others may use the data long after the creator has finished with it and is available to answer questions.
4. Design of the data base should be a slow, careful affair. If this stage is neglected, the data base designer risks wasting many hours of work and compromising the real value of the data base.

E. RECOMMENDATIONS FOR IMPLEMENTING A LARGE DATA BASE

Statgraphics has a scrollable data editor which allows the entry, manipulation, and review of large data bases. It is convenient, simple to use, and, most importantly, makes it easy to correct and manipulate the data when anomalies are detected.

In view of the value that such a scrollable data editor provided when reducing and documenting a data base which is already in existence, here are some recommendations for data base design. The design should:

1. Allow for speedy input of and access to new data;
2. Allow the data to be manipulated and massaged with scrollable full-screen data editors;
3. Allow easy access by statistical graphics packages such as Grafstat.

III. DATA ANALYSIS

A. GENERAL APPROACH FOR THE EXPLORATORY DATA ANALYSIS

The primary question which this thesis attempts to answer is, "What relationships exist between the information derived from the subjects' background investigations and the final disposition of their service?" The answers obtained here will not, of course, be all inclusive but provide a starting point for further research involving this data base. In particular, this is not the only question to be answered from the data base, but as in much research, other questions and facts become apparent as the research progresses.

Inherent in a data analysis is the initial investigation into the properties and limitations of the data. The PERSEREC special background investigation data (SBID) is primarily categorical in nature. The record for each individual contains several different types of information:

1. Background and biographical information such as age, marital status, reason for investigation, etc.;
2. Derogatory information (or lack thereof) obtained by investigators from various sources (high school, neighborhood, employers, etc.); this information may consist of crimes, subject admissions, and other matters that reflect on the person's character and judgement;
3. Recommendations from various people associated with these sources as to whether they felt that the individual in question should be trusted with a position of trust and responsibility;
4. The result of the term of military service, whether the individual was discharged normally, or due to some adverse circumstances.

The data can viewed as information obtained prior to the completion of the investigation (explanatory variables or independent variables) and information which is the result of the person's service after the investigation (response variables or dependent variables).

Note that the data is basically categorical, e.g., male or female, and thus has no inherent ordering. Thus, while frequency counts can be obtained and are given in Appendix A, no distributional measures, e.g., means or variances, can be computed. Similarly, dependencies and associations cannot be measured by moments based upon joint distributions, e.g., correlation coefficients.

The data thus appears to be ideal for contingency table methods [Ref. 3 : pp. 153 - 170]. However, note that the one response variable in the contingency table is almost

always the character-of-service categories (six of them), and these have fixed marginal frequencies of occurence. Thus the type of contingency table analysis applied is for differences in probabilities (of the occurrences of the free category in each of the fixed category samples). Thus, one would test that the probability of discharge for each of the six categories (good, homosexual, misconduct, drug/alcohol abuse, court martial, character/behaviour disorder) is the same for males or females.

The background, derogatory, and recommendation data is very specific and limited to that obtained from each type of source. The information received from any one source, say an individual's high school, does not paint a very complete picture of the person, no matter whether that information is primarily good or bad. Of more interest is the overall characterization of the individual's past which is represented by the record as a whole. The obvious question is how can many pieces of different information, i.e., different columns in the data base, be combined to give an overall picture of the person's character?

There are several approaches used in this thesis to answer that question:

1. To look at particular biographical information which may provide insight into discharge categories;
2. To look at the total quantity of derogatory information as a measure relating to the discharge categories;
3. To look at the existence and quantity of particular types of derogatory information as a measure relating to the discharge categories, i.e., evidence of drug use discovered in the investigation;
4. To look at the types, amounts, and quality of the recommendations that persons were willing to give about the subject of the investigation.

The data used for this research was configured as coded APL vectors ordered by record number such that all information for record number 1 was in the first position of each variable vector and so forth for all records. The numeric codes represented each particular category of the information contained in that type of variable vector.

It was a relatively straightforward matter to use APL logical operations to collect the information needed for each approach discussed above and then to represent this information in an APL vector. For example, if the existence of drug use in the records was the information desired, each of the 35 columns which represented derogatory information were operated upon in turn and a 0 or 1 resulted which represented no drug use or drug use in a particular column. As this information was obtained the information was totalled so that in the end a vector representing the total amount of drug use

discovered by the investigators was available. The result was an ordered, integer-valued APL vector which could be cross-tabulated with the character-of-service vector to determine if the character-of-service was independent of evidence of prior drug use, e.g., the probability of discharge for drug/alcohol abuse was the same for any level of prior drug abuse. With an easy APL logical operation this variable could be quickly transformed into a binary vector with 0 representing no evidence of drug use and 1 representing at least one incident on record.

The information, once configured into appropriate independent variable vectors, and the response variable vector (generally character-of-service) were cross-tabulated to create a contingency table, generate residuals, and produce the chi-square test statistic allowing a test of difference of probabilities to be performed. The graphical mainframe software package, Grafstat, did not offer a cross-tabulation capability; however, a customized APL cross-tabulation function developed by Luis Uribe and Professor Peter A. W. Lewis was used which operated in the Grafstat environment. This package was developed to allow, simultaneously, a visual and tabular way of looking at the cross-tabulation of the categorical data while also providing the needed information to quickly perform any of the chi-square contingency tests (the independence test, the difference of probability test, or the fixed marginal total test). These tests are conducted in identical manners; however, they differ in the manner in which the data is sampled and thus in the way the test is interpreted. Figure 4 shows an example of the output of the cross-tab function.

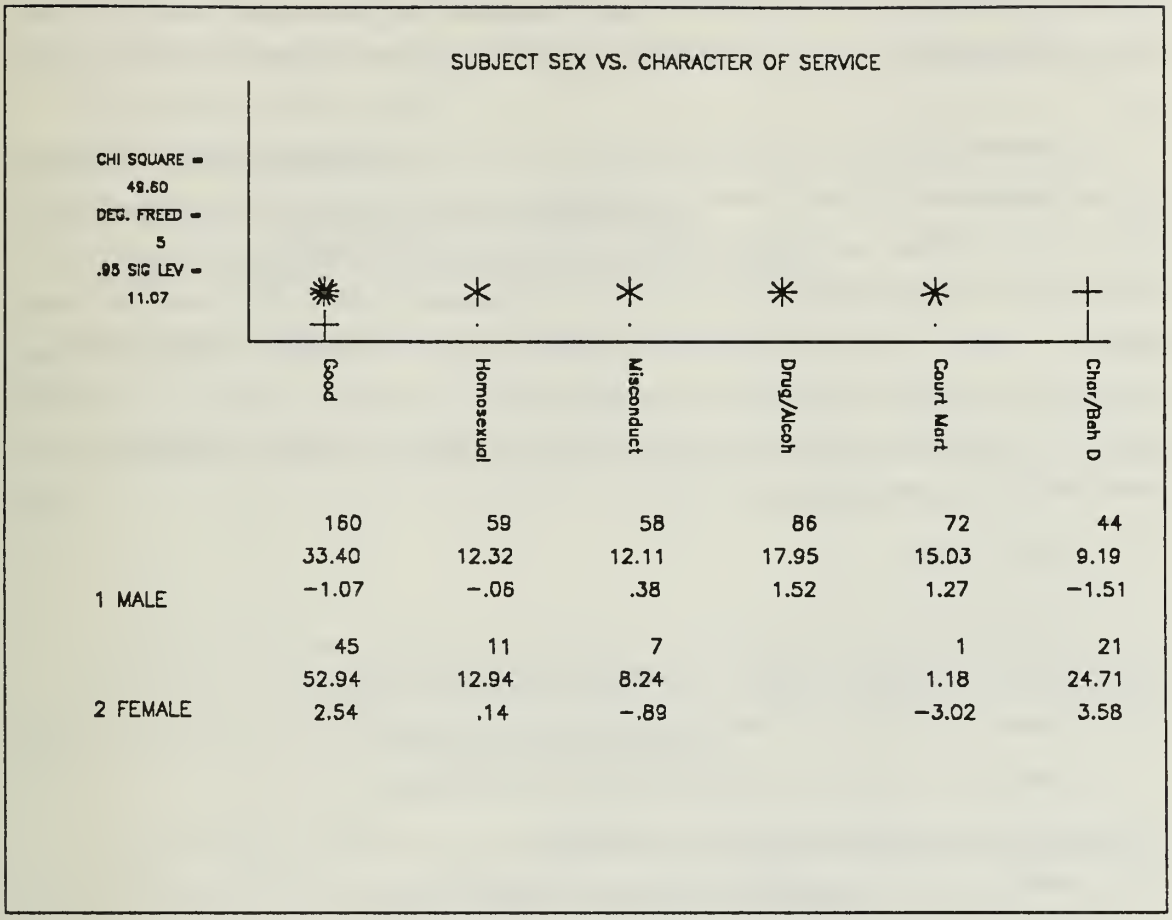


Figure 4. Sample Cross-Tabulation: The Cross-Tabulation of Sex vs. Character of Service.

The graphical display in Figure 4 is a combination of tabular data representing the resultant contingency tables of the cross-tabulation and scaled sunflower plots representing the relative counts produced in the table. The sunflower plots are coded in the same position and color (on the color graphics screen) as the related tabular information. The number of arms in the sunflower plot is proportional to the total counts in the table. The graph is designed so as to make it easy for a user to quickly locate cells with high (or low) incidence. Thus one can quickly see from the graph in Figure 4 that the highest incidence is of males having good discharges and that no females in the sample were discharged for drug/alcohol abuse.

The table below the graph gives the actual count for each cell, the cell's percentage of the particular row marginal total, and the residual produced using the standardized

difference between the actual cell count and the expected cell count (see below for definition). When color output is available, 'significantly' large residuals are displayed in red.

The cross-tabulation is a simple procedure. Two vectors representing two different types of information on each person are configured with a numerical code representing specific classes within each vector. The vectors are ordered identically so that the information in the first position of both vectors represent information about the same person (or record) and so forth for all record positions. The cross-tabulation is merely a count of the various matchups of categories between the vectors. Figure 4 shows the cross-tabulation of the subjects' sex versus their character-of-service. The variable representing sex is coded as:

0 - represents male.

1 - represents female.

The variable representing character-of-service is likewise coded as:

0 - represents a discharge for service characterized as good.

1 - represents a discharge for homosexual-related problems.

2 - represents a discharge related to misconduct.

3 - represents a discharge related to drug or alcohol abuse.

4 - represents a discharge as a result of a court martial.

5 - represents a discharge for a character or behaviour disorder.

The character-of-service categories are the fixed-size categories (columns in Figure 4). The sex of the individuals represent the free-size categories (rows in the same figure).

The upper left cell in Figure 4 shows a count of 160 where the row designated "male" and the column designated as "good" intersect. This occurred because there were 160 records where the code 0 representing sex matched up with the code 0 representing character-of-service. Each cell in the cross-tabulation was constructed in this same manner.

The cross-tabulation function used in this thesis assisted in the conduct of the chi-square test of differences in probability. This test is conducted when separate samples are taken from several populations. In this case the several populations are the six populations of people with different character-of-service records, these populations having

been fixed at the beginning of the data collection. There are n_j observations possible from each separate population, j and $N = n_1 + n_2 + n_3 + n_4 + n_5 + n_6$ total observations.

Assumptions of the test are:

1. Each sample is random.
2. Outcomes of various samples are independent.
3. Each observation can be classified into one of m classes.

First here is a definition of terms to be used. Let,

o_{ij} = the cell count in the cell formed from row i and column j, i.e., o_{12}, in Figure 4, is 59, the number of males (i = 1) with homosexual 'character-of-service' (j = 2).

r_i = the marginal total of counts which fall in the category represented by category i, which represents the total of all entries from row i. Thus, in Figure 4, r_1 is the total number of males and r_2 is the total number of females.

n_j = the marginal total of counts which fall in the category represented by category j, which represents the size of the sample from the j^{th} population.

$N = \sum_{j=1}^{n} n_j$; the total number of records.

$i = 1, 2, ..., m$; the index of the random row categories.

$j = 1, 2, ..., n$; the index of the fixed column categories.

p_{ij} = the probability that an individual possesses classification i, given that they are a member of sample j. In the example of Figure 4 this is the probability that a person is male or female (classification i) given that the person is from a particular character-of-service category.

The hypotheses of the chi-square test of differences in probabilities are (if the column sums are the fixed categories):

H_0 : All probabilities in the same row are equal to each other ($p_{i1} = p_{i2} = \cdots = p_{in}$) for all i.

H_1: : At least two probabilities are not equal to one another ($p_{ij} \neq p_{ik}$ for some i and some pair j and k). [Ref. 3 : p. 154].

In other words, in Figure 4 , this hypothesis suggests that the probability that a person in the good character-of-service category is male is equal to the probability that a person of the misconduct character-of-service category is male and so forth for the other categories.

Alternatively, the null hypothesis, H_0, is defined as:

H_0: Pr [an observation falls in cell i,j]

= Pr[an observation falls in row i] × Pr [an observation falls in column j]

for all i, j.

The expected count in each cell, under the null hypothesis of equal probability, can be defined as follows:

$$e_{ij} = \frac{r_i \times n_j}{N} .$$

The standardized difference between the actual cell count and the expected cell count is called the signed-residual and is calculated in the following manner:

$$\frac{o_{ij} - e_{ij}}{\sqrt{e_{ij}}} .$$

The test statistic, Q, is determined by summing the squared residuals:

$$Q = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} .$$

Q is, for large enough cell entries, distributed as a chi-square statistic with $(m - 1) \times (n - 1)$ degrees of freedom. The chi-square statistic Q, it's degrees of freedom,

and the 95th quantile of the matching chi-squared distribution are listed in Figure 4 to the left of the graph to aid in evaluating the hypotheses.

The manner in which the cross-tabulated information was displayed allowed the chi-square test of independence to be performed in a relatively simple manner. The test required only a quick comparison of the test statistic and the 95th quantile to evaluate the null hypothesis that the probabilities in each of the same row were equal.

There remained an evaluation of the expected cell sizes for appropriate size for the chi-square distribution of the test statistic to be valid and for the test to be meaningful. The generally accepted rule that the expected value of each cell, e_{ij} should be larger than 5.0 was followed, and if this were not so the number of categories should be reduced by combining like categories. There seems to be some support for an argument that, given large enough samples, an expected cell size even as small as one is acceptable. However, the purpose of this thesis is not to argue this point, and therefore the rule was followed. [Ref. 3 : p. 156]

At this point in the analysis, if expected cell sizes were too small, like rows were combined and the cross-tabulation was reiterated, or, if expected cell sizes were large enough, the cross-tabulation was accepted as performed.

B. RESIDUAL ANALYSIS

The chi-square test of independence conducted with a contingency table is very general and does not say anything specific relating to the likely dependencies of particular categories within the variable vector.

For example, incidents of prior drug use, cross-tabulated with character-of-service may indicate a lack of independence; however, it is difficult to say, specifically, that lack of prior drug use indicates a higher than expected probability of good service, or that high levels of prior drug use indicates a lower than expected probability of good service, or both, even when the chi-square test of differences in probabilities indicates that the probabilities that a member of a fixed category is classified in the each of the free categories is not equal. The problem becomes more severe with increase in the number of categories, though for a 2 by 2 table there are clear constraints because a large number in one cell forces the other cell to be small. Thus an empirical method of determining components of dependence information is desired.

The residuals produced within each cell represent a rough way of determining such dependency: a higher magnitude of the residual seems to indicate a larger discrepancy from the expected value of the cell size given independence. If the sample is large

enough, the signed residual is approximately normally distributed. Thus we can score all residuals and select those whose values are greater than, say, the 95th quantile of the standard normal distribution. However, we cannot make this assumption when dealing with more than one residual produced from two dependent variables. It is difficult to say much about the distribution of the maximum residual produced from two variable vectors even under the null hypothesis of independence.

The basis of the analysis presented in this thesis is to look at just the sort of dependency mentioned above and some way was sought to analyze the residuals to provide an acceptable basis for asserting that the particular dependencies between variables existed.

Residual size appeared to be the key to determining the particular dependencies within variables which have been shown to not be independent. Ideally, all residuals would be distributed according to some known distribution, and a residual with an unlikely value under this distribution would indicate a particular dependency. There is no basis for such an ideal situation. A technique which is available, however, can provide an empirical method to give insight into this distribution. This method is known as the bootstrap-simulation technique.

The premise behind the bootstrapping technique is that there exist many situations where it is desirable to have distributional information about random variables where such information is unknown or difficult to generalize. In cases where such information is unknown but the distribution of the underlying basis for the random process is known, then to simulate the process and sample the random variables created allows an empirical analysis to be conducted, especially of the quantiles of the unknown distribution.

The basis for the bootstrapping used in this thesis is that independent variables produce counts in the contingency tables according to the null hypothesis explained in the previous section. That is, the probability that a random variable belongs in a particular cell (i,j) is equal to the probability that it belongs in the particular row category i times the probability that it belongs in the particular column category j. Taking this hypothesis a step further, the probability that a count belongs in a particular cell of a contingency table formed by two independent variables can be estimated by dividing each cell's expected value by the total counts contained in the table. The total of all the estimated probabilities sum to one and the distribution of counts within the table are essentially multinomial with n total counts and the probabilities of each cell.

With this in mind, a monte-carlo simulation of any particular contingency table (with particular marginal counts) can be performed using multinomial random numbers. The distribution of the largest absolute residual can be empirically generated with a large number of replications. This allows the comparison of the 95th quantile residual generated by this simulation of independent variables versus the residuals produced in an actual contingency table, where the hypothesis of independence has been rejected.

In effect, the comparison of the residuals against the 95th quantile bootstrap residual is a test of the null hypothesis:

H_0 : The residual does not indicate a direct relationship between the two factors which join to form a cell.

If the magnitude of a particular residual is larger than the bootstrap residual, then we reject this H_0 at a level of significance of 0.05.

For this thesis a FORTRAN program was written which used as input the actual marginal totals produced in each cross- tabulation performed. These marginal totals were used to generate the expected values for each cell under the null hypothesis that the variables were independent. Probabilities for each cell were computed. The IMSL multinomial random number generator was used to generate counts based upon these probabilities and the total count for the table. The simulation was replicated 200 times for each table. Residuals were generated for each cell in the exact same manner as used in the cross-tabulation function for each replication. The largest absolute residual for each replication was determined and saved. An empirical distribution of the largest residuals of a particular m by n contingency table with particular marginal totals was available. The statistics of interest (in particular,upper quantiles) were available for comparison with the residuals produced by the cross- tabulation of the actual variable vectors. Appendix C contains the FORTRAN program, Appendix D contains the input file, and Appendix E contains the output file.

Comparison of the actual residuals with the 95th quantile largest residual produced in the simulation allowed a determination of the comparative size of residuals and allows some insight into particular dependencies within each contingency table.

C. APPLICATION TO THE ACTUAL DATA

There were two major types of independent variables available from the data: derogatory information consisting of crimes and negative items of information uncovered

by investigators, and recommendations from friends, family members and acquaintances of the subject.

There were two primary response variables available in the data base: character-of-service and type of discharge. The character-of-service represents the manner in which the subject performed during his period of service. If the individual completed his enlistment without incident, his service is characterized as good, otherwise the individual has an adverse entry in his file and was released from service early. The categories of service are:

1. Good.
2. Homosexual activity or inclination.
3. Misconduct.
4. Drug or alcohol abuse.
5. Court martial.
6. Character or behavior disorder.

Type of discharge (honorable, general, and other than honorable) is highly dependent upon the character-of-service. A subject with a negative character-of-service category sometimes receives an honorable discharge, while a subject with a good character of service never receives anything but an honorable discharge. For this reason, the character-of-service was used in this thesis as the response variable in all cases. In fact, because the records were selected based upon the character of service categories, the character-of-service would appear to be the only reasonable response variable.

D. ANALYSIS OF DEROGATORY INFORMATION

1. General

Derogatory information uncovered by the investigators came from the high schools, colleges, employers, neighbors, national agencies (FBI and police departments), local agencies, credit bureaus and other sources ('other sources' represents sources of information exclusive of the other, explicitly-stated sources). The code or codes representing derogatory information obtained from each of the source categories for each investigation subject was listed in the data base in the appropriate column variable vector.

2. Tests for Differences in Probability Involving Derogatory Information

The derogatory information of interest was grouped and collated into a new, coded variable vector and cross-tabulated with the character-of-service.

Groupings were determined to confirm relationships which should exist (record of prior drug use might indicate a higher than normal result of drug and alcohol character-of-service, for example) and also to answer some specific questions PERSEREC was interested in investigating.

Table 4 summarizes the results of the cross-tabulations and tests for independence concerning derogatory information (the response variable in all cases is the character-of-service), which are shown in later parts of this section.

Table 4. SUMMARIZED RESULTS OF CROSS-TABULATIONS: Involving derogatory and other information.

Independent Variable	Chi-Square Statistic	Degrees of Freedom	.95 Level of Significance	Dependency?
Sex (male female)	49.6	5	11.07	yes
High School Diploma	18.07	5	11.07	yes
Age at Enlistment	2.63	10	18.31	No
Derogatory Info. Disclosed	29.81	5	11.07	yes
Incident of Major Crime	35.07	5	11.07	yes
Adjustment Incidents	17.07	5	11.07	yes
Prior Drug Use	39.24	5	11.07	yes
Prior Drug Alcohol Use	38.64	5	11.07	yes
Category of Most Serious Drug Use	60.46	10	18.31	yes

As you can see, all of the cross-tabulations listed above indicate a difference in structure between character-of-service categories except for the age at enlistment-character-of-service comparison.

3. Residual Analysis Involving Derogatory Information

Residual analysis of the cross-tabulations is explained in the following tables. Negative residual values indicate that the actual cell count was lower than the expected value. This indicates a negative relationship. The opposite holds for positive residuals.

As displayed in Table 5 , the incidence of court martials is significantly low among women (the second largest residual in absolute value, -3.02), but character and

behavior disorders are significantly more prevalent among women (largest residual in absolute value).

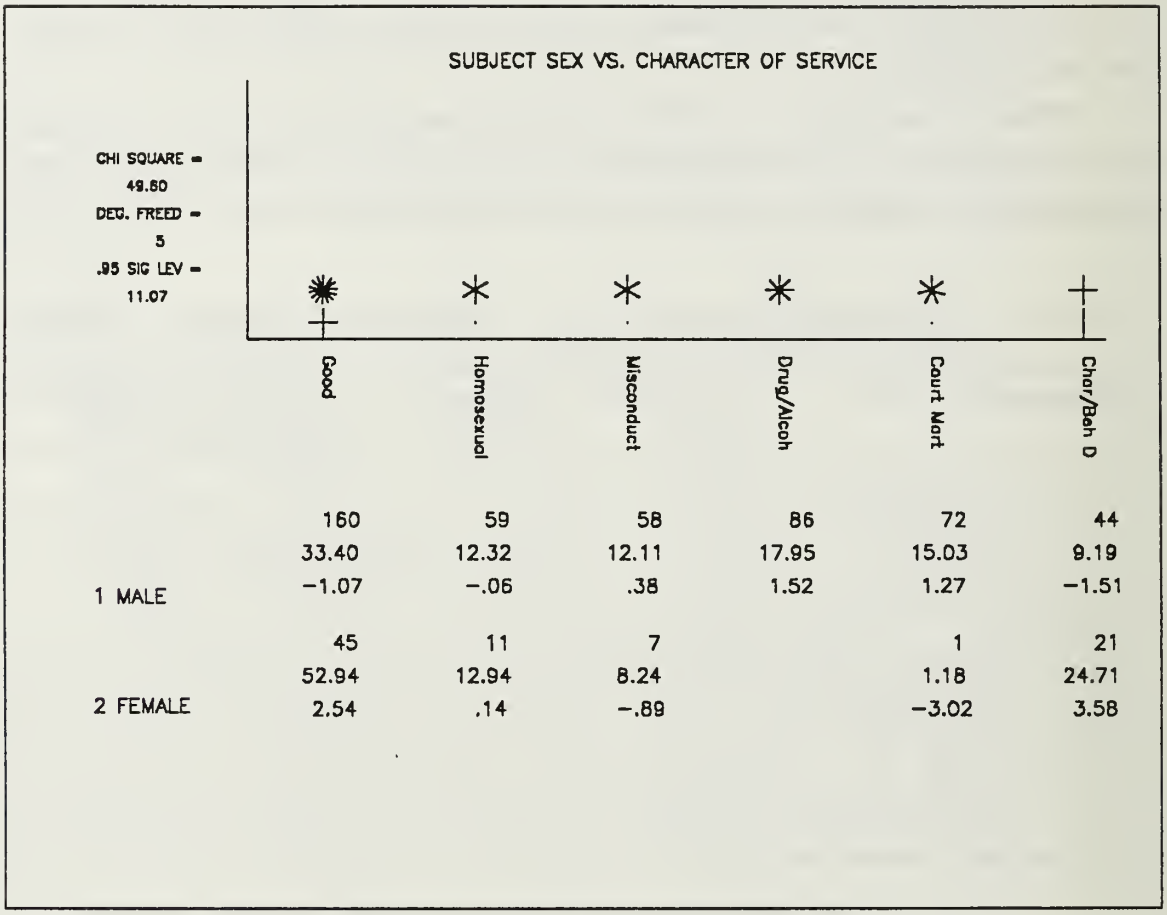


Figure 5. The Cross-Tabulation of Sex vs. Character-of-Service: This is a another copy of Figure 4 .

Table 5. RESIDUAL ANALYSIS OF SEX VS. CHARACTER-OF-SERVICE: For Figure 5

Independent Variable	.95 Quantile Bootstrap Residual Value	Significant Categorical Relationship	Residual Value
Sex	2.71	Female - Court Martial	-3.02
		Female - Character Behavior	3.58

Table 6 shows that those individuals receiving court martials and being discharged have a significantly higher incidence of being non-graduates of high school.

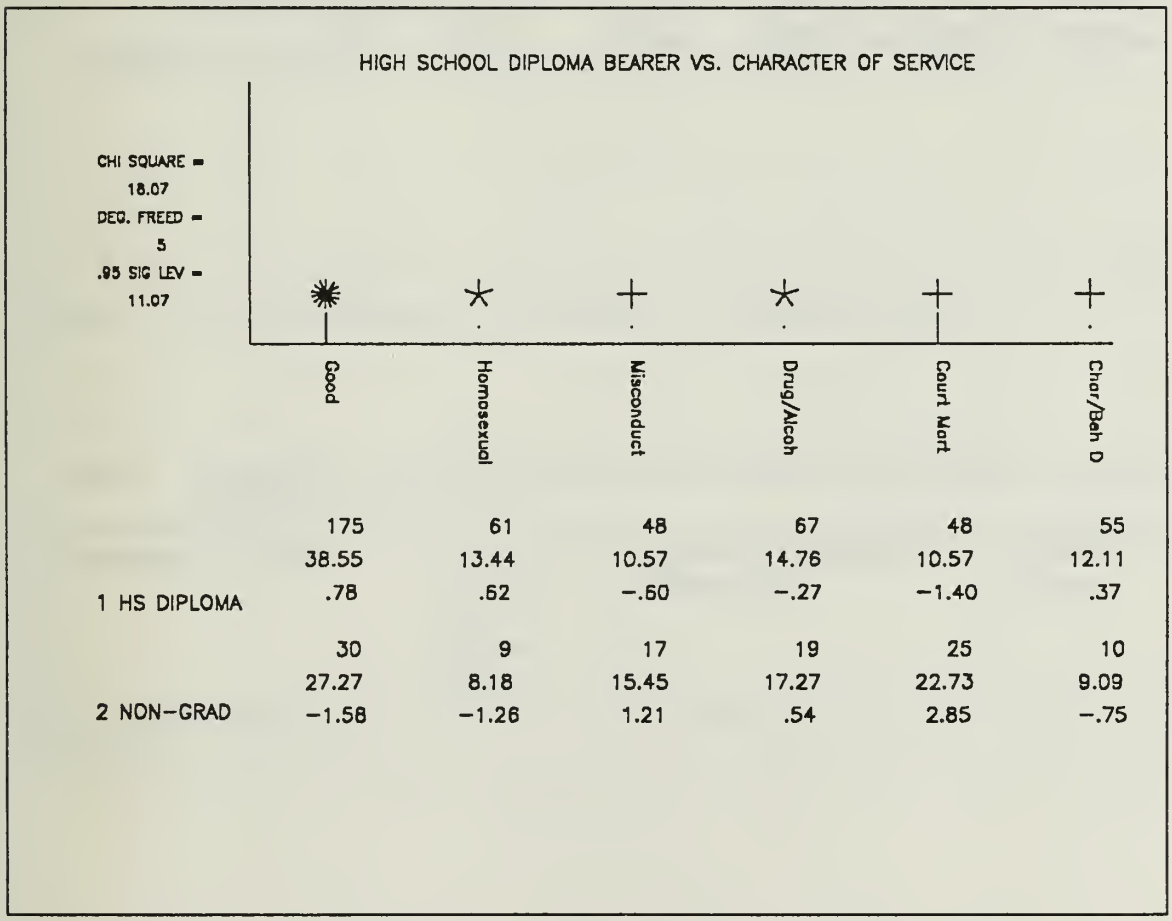


Figure 6. The Cross-Tabulation of High School vs. Character-of-Service.

Table 6. RESIDUAL ANALYSIS OF HIGH SCHOOL VS. CHARACTER- OF-SERVICE: For Figure 6

Independent Variable	.95 Quantile Bootstrap Residual Value	Significant Categorical Relationship	Residual Value
High School Diploma	2.69	Nongraduate - Court Martial	2.85

As intuition, and Table 7 suggests, those individuals who have no derogatory information disclosed in their investigation were significantly more likely to have successfully completed their service (received a good character-of-service designation).

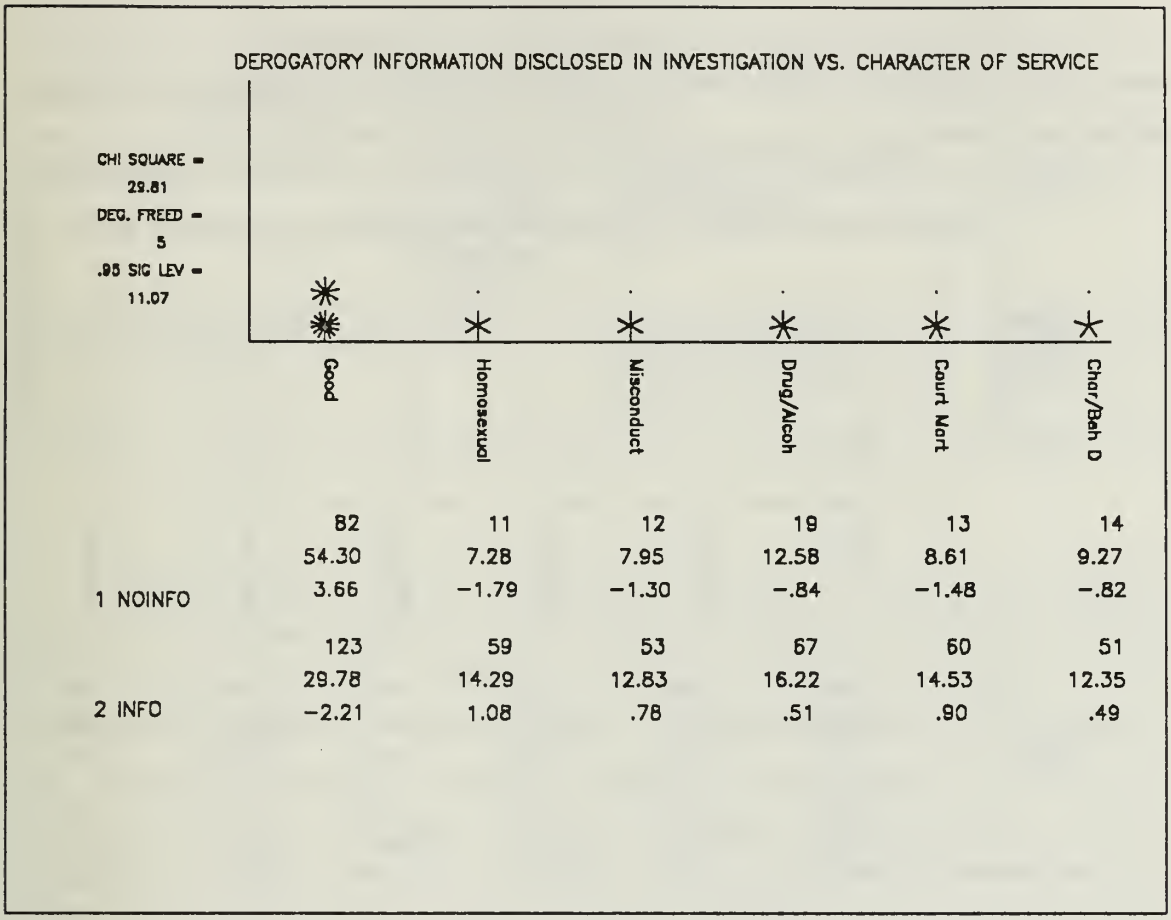


Figure 7. The Cross-Tabulation of Derogatory vs. Character-of-Service.

Table 7. RESIDUAL ANALYSIS OF DEROGATORY VS. CHARACTER- OF-SERVICE: For Figure 7

Independent Variable	.95 Quantile Bootstrap Residual Value	Significant Categorical Relationship	Residual Value
Derogatory Information Disclosed	2.90	No Information - Good	3.66

Looking at Figure 8 we see that persons who had a record of major crime (felony-related derogatory information) were more likely to be discharged for courts martial while persons with a major crime on record were less likely to be discharged with a good character-of-service.

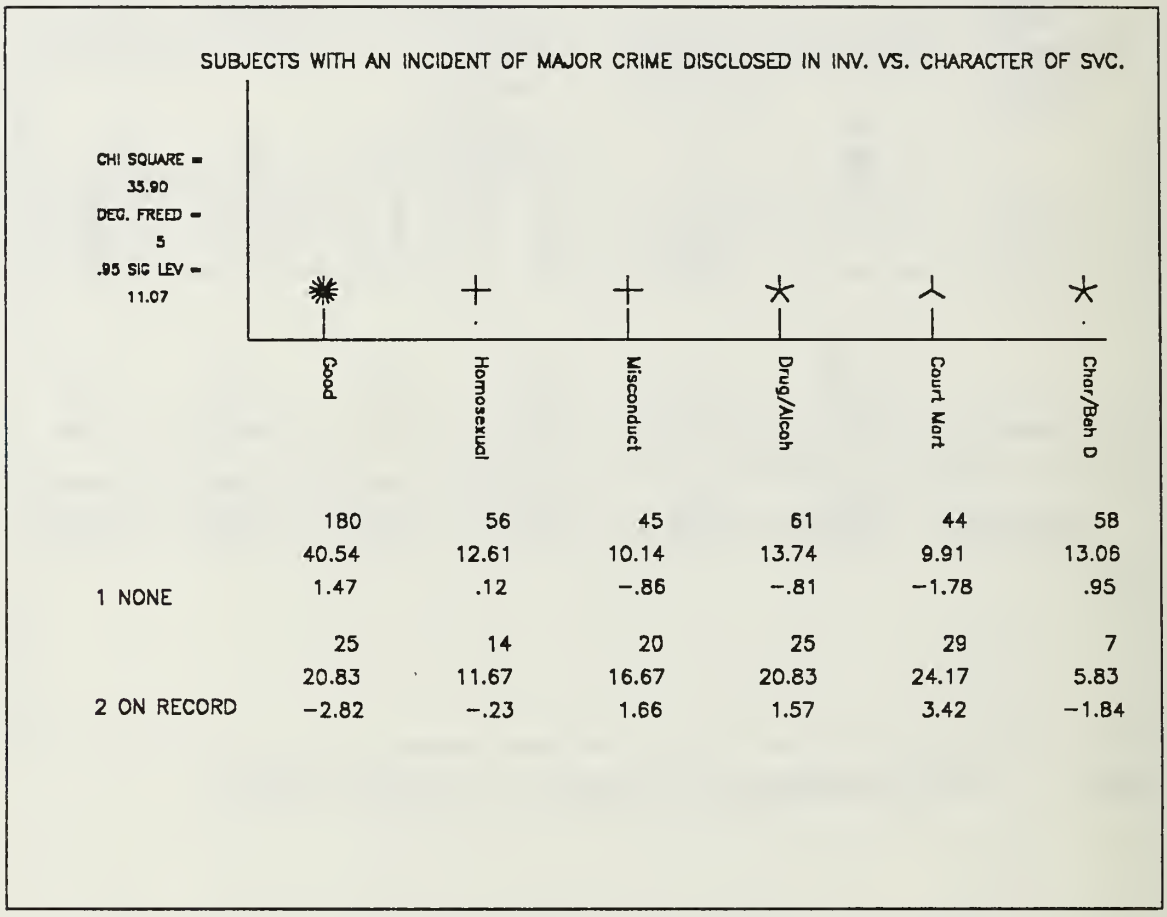


Figure 8. The Cross-Tabulation of Major Crime vs. Character-of-Service.

Table 8. RESIDUAL ANALYSIS OF MAJOR CRIME VS. CHARACTER- OF-SERVICE: For Figure 8 on page 31

Independent Variable	0.95 Quantile Bootstrap Residual Value	Significant Categorical Relationship	Residual Value
Major Crime Disclosed in the Investigation	2.84	On Record - Good	-2.82
		On Record - Court Martial	3.42

Table 9, Figure 9, Figure 10, and Figure 11 all refer to drug abuse. Figure 9 considers the incidence of drug use alone. Figure 10 is concerned with substance abuse in general by including alcohol-related incidents along with drug use. Figure 11 looks at drug use by categories of seriousness, from none, to marijuana only, to any involvement of a more serious nature. Evidence of drug/alcohol abuse prior to the investigation is highly associated with the drug/alcohol character-of-service (Figure 10). Records with no evidence of prior use were more likely to result in the good character-of-service. Combining the prior alcohol incidents did not significantly alter this relationship (Figure 11). There was a strong relationship between those court martialed and a record of hard drug use (drugs stronger than marijuana), shown in Figure 11. Prior evidence of marijuana use is less likely to result in a good character-of-service (Figure 11).

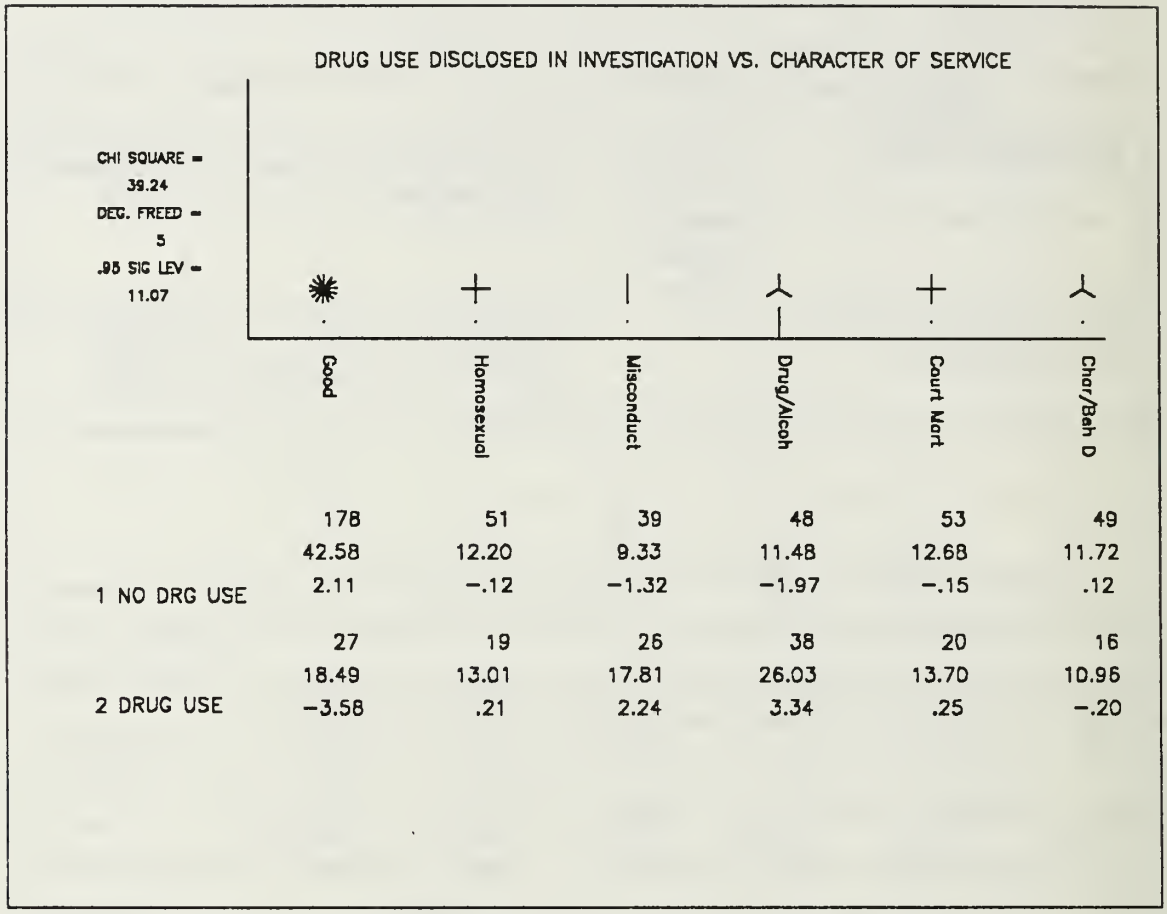


Figure 9. The Cross-Tabulation of Drug Abuse vs. Character-of-Service.

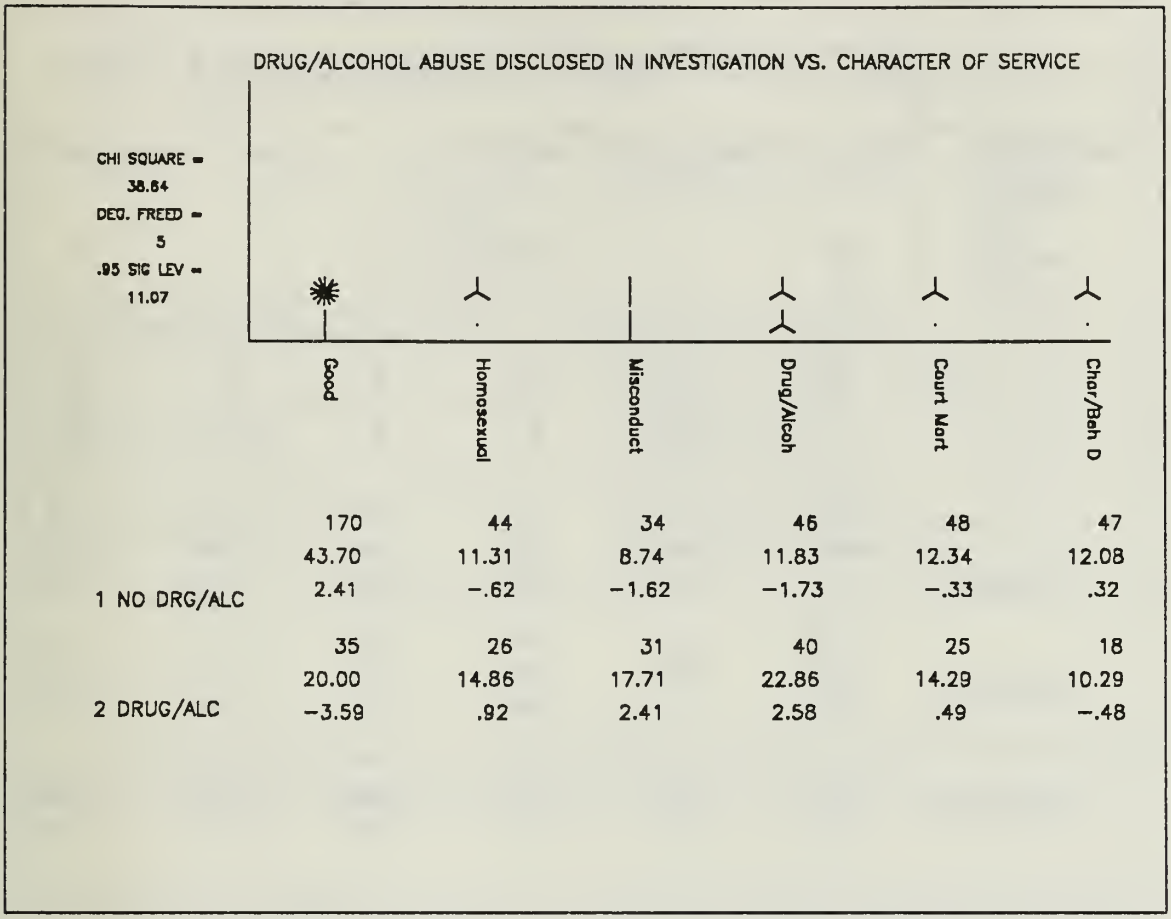


Figure 10. The Cross-Tabulation of Drug/Alc. Abuse vs. Character-of-Service.

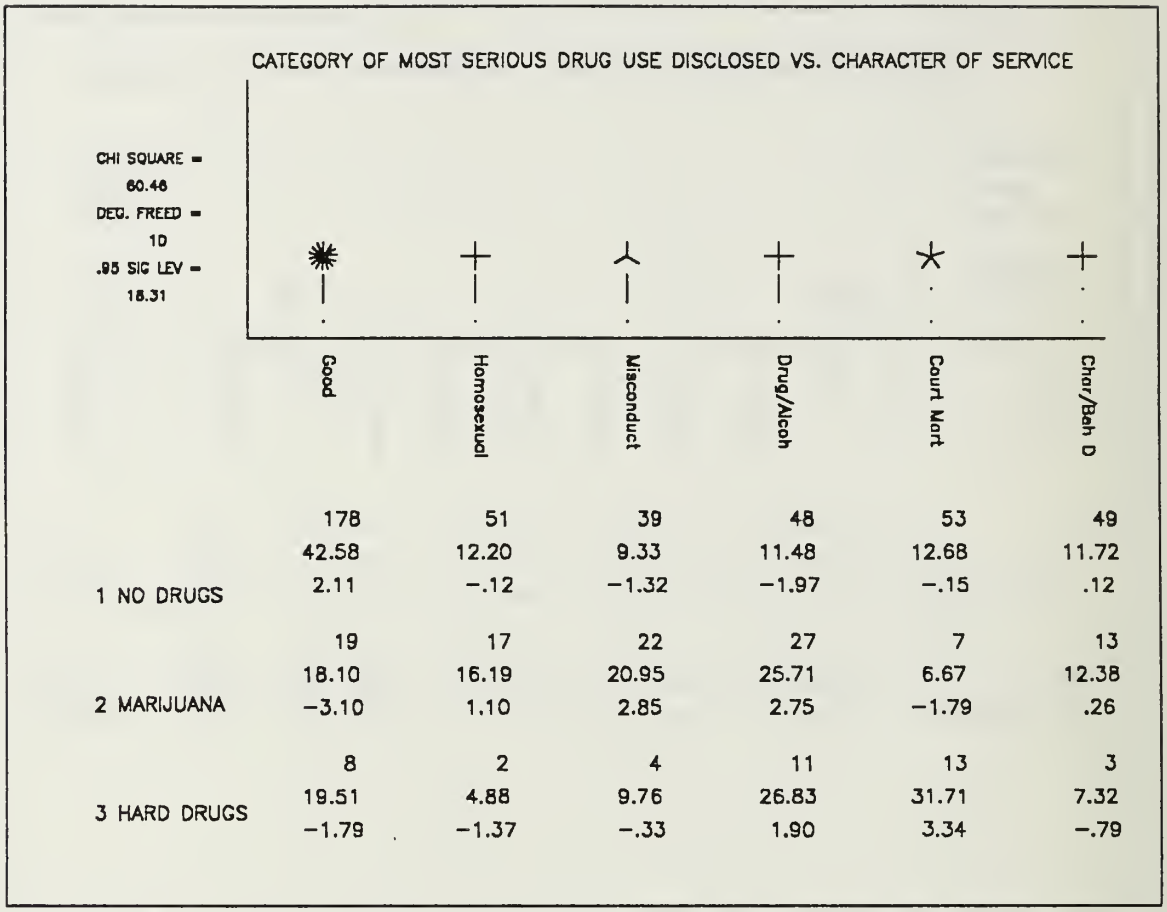


Figure 11. The Cross-Tabulation of Most Serious Drug Abuse vs. Char.-of-Svc.

Table 9. RESIDUAL ANALYSIS OF DRUG USE VS. CHARACTER-OF- SERVICE

Independent Variable	0.95 Quantile Bootstrap Residual Value	Significant Categorical Relationship	Residual Value
Drug Use Disclosed (Figure 9)	2.57	Drug Use - Good	-3.58
		Drug Use - Drug/Alcohol	3.34
Drug/Alcohol Abuse Disclosed (Figure 10)	2.56	Drug/Alc - Good	-3.59
		Drug/Alc - Drug/Alc	2.58
Category of Most Serious Drug Use (Figure 11)	2.91	Marijuana - Good	-3.10
		Hard Drugs - Court Martial	3.34

The cross-tabulation of age versus character-of-service shown in Figure 12 indicated no separate association between the various character of service categories.

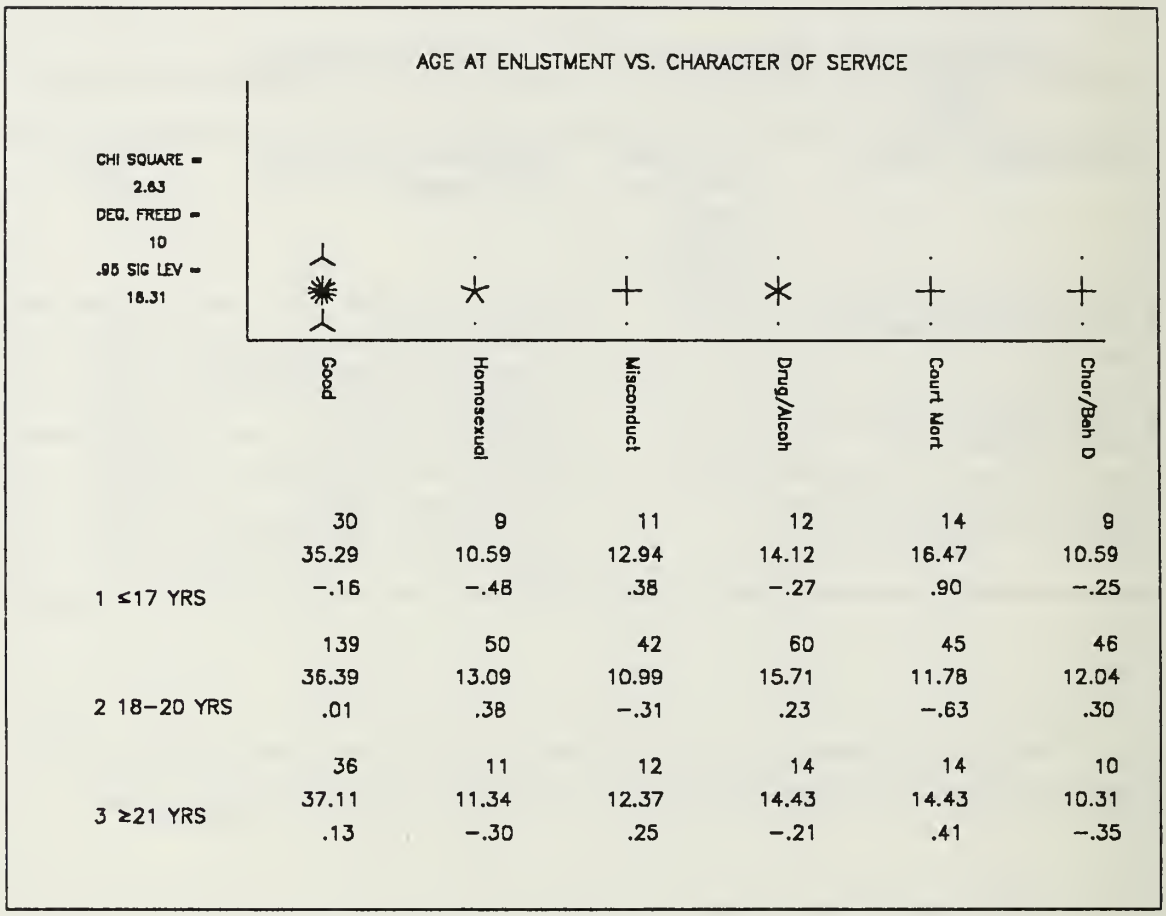


Figure 12. The Cross-Tabulations of Age Versus Character-of-Service

4. General Comments About the Derogatory Information Cross-Tabulation

There are some general comments which follow from the analysis of the derogatory information cross-tabulations displayed in Figures 5 through 12. Significantly lower than expected amounts of derogatory information were associated with the good character-of-service category in almost any cross-tabulation performed. The residual value was not always large enough to exceed the 95th quantile bootstrap residual; however, it is significant that persons with relatively low amounts of derogatory information are almost always associated with the good character-of-service. Some amounts of certain types of information are more strongly associated with some categories than others. But some "good" people had "bad" information in their files and some "bad" people had no "bad" information in their file. All that can really be inferred is trends and tendencies, not predictions. There does appear to be plenty of evidence to indicate caution should

be exercised when investigations indicate serious derogatory information. The appearance of derogatory information does not guarantee a problem service member, however, the government is indeed taking a higher risk that the service of such a person will terminate early. The mere existence of derogatory information does not indicate that a service member will leave service early (roughly 40% of subjects with good character-of-service had at least one item revealed in their investigation); but, of those that left service with an adverse discharge, 80% had at least one item of derogatory information.

E. ANALYSIS OF RECOMMENDATION DATA

The investigators interview persons who know the subject of an investigation and ask them:

1. Would you recommend this person for a position of trust?
2. Or, would you recommend that person for a position of trust with supervision?
3. Or, would you not recommend them?
4. Or, do you decline to comment?

This information is used, along with everything else, to determine if the person gets the security clearance. Questions that came to mind when the data was investigated were:

1. "What do positive and negative recommendations tell us in relation to the subsequent character-of-service?
2. Can they tell us something separately?
3. Can they be combined, by some scoring system, to allow us to show some relationship between the recommendations received (both positive and negative) and the subsequent character-of-service?"

The incidence of negative recommendations is relatively rare in this data base (only 73 persons had at least one "not recommended", and few of them had more than one).

The recommendation information was cross-tabulated against the character-of-service variable in several ways:

1. A total of the number of recommendations each individual received was made (case 1).
2. A total of the number of "not recommended" each individual received was made (case 2).
3. A recommendation score was constructed with a recommendation having a +1 value and all other recommendation categories having a -1 value (case 3), called the adjusted recommendation score.

4. A recommendation score was constructed with a recommendation having a +1 value, recommendation with supervision and decline comment having a -1 value, and not recommended having a -3 value (case 4), called the weighted recommendation score.

Figure 13, Figure 14, Figure 15, and Figure 16 below show the cross-tabulations of the four recommendation cases discussed in the previous paragraph versus the character-of-service variables.

The cross-tabulations of the four recommendation scoring methods with the character-of-service are summarized in Table 10 .

Table 10 shows, for all cases, that the hypothesis of equal probabilities for the recommendation levels between the character-of-service categories should be strongly rejected.

NUMBER OF RECOMMENDATIONS DISCLOSED IN INVESTIGATION VS. CHARACTER-OF-SERVICE

CHI SQUARE = 82.89
DEG. FREED = 15
.95 SIG LEV = 25.00

	Good	Homosexual	Misconduct	Drug/Alcoh	Court Mart	Char/Beh D
1 NO RECS	1 1.61 −4.54	9 14.52 .47	19 30.65 4.43	18 29.03 2.78	6 9.68 −.71	9 14.52 .69
2 1−3 RECS	25 29.76 −1.00	10 11.90 −.13	9 10.71 −.22	15 17.86 .61	13 15.48 .65	12 14.29 .75
3 4−7 RECS	143 45.54 2.70	27 8.60 −1.92	28 8.92 −1.36	40 12.74 −1.14	47 14.97 1.00	29 9.24 −1.19
4 ≥8 RECS	36 34.62 −.29	24 23.08 3.09	9 8.65 −.86	13 12.50 −.72	7 6.73 −1.76	15 14.42 .87

Figure 13. The Cross-Tabulation of Recommendations vs. Character-of-Service (Case 1).

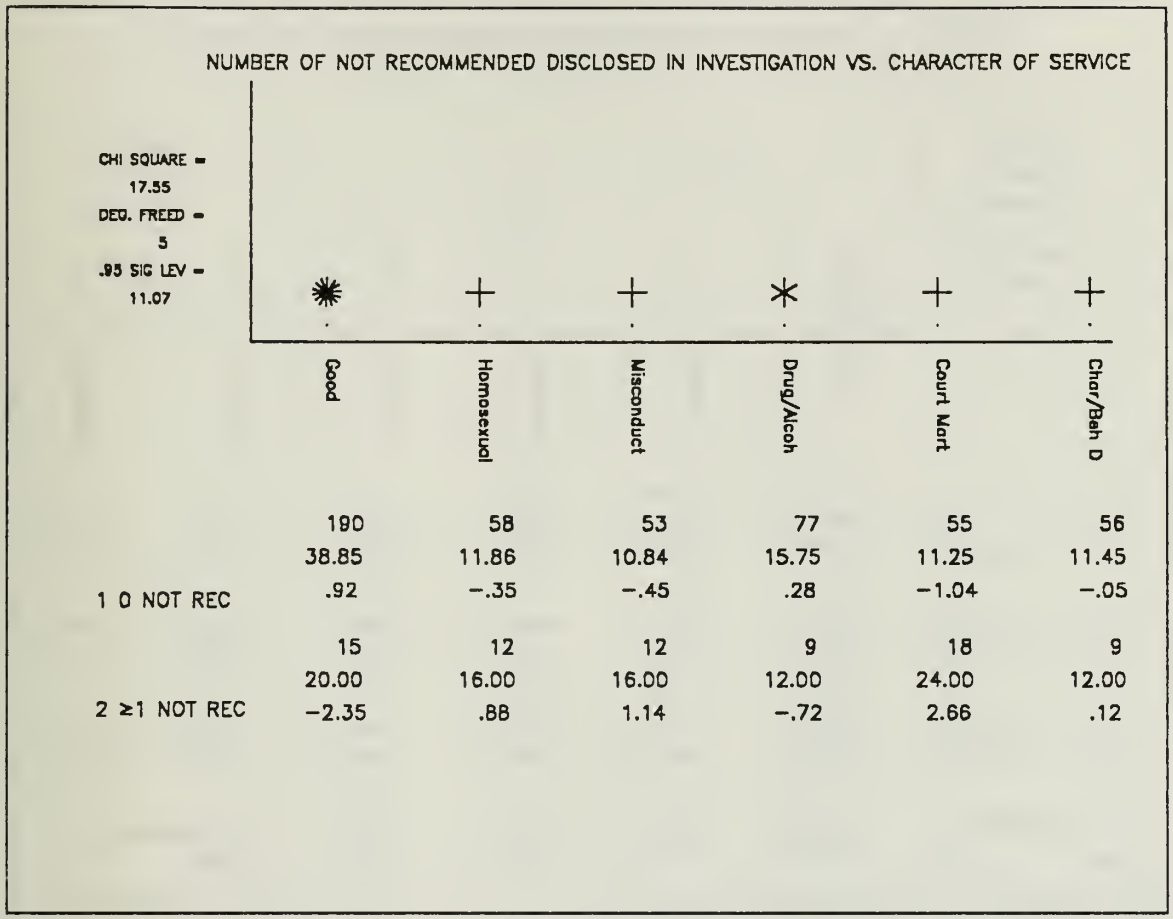
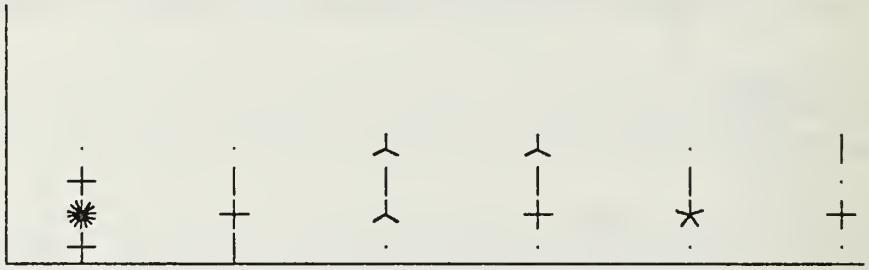


Figure 14. The Cross-Tabulation of Not Recommended vs. Character-of-Service (Case 2).

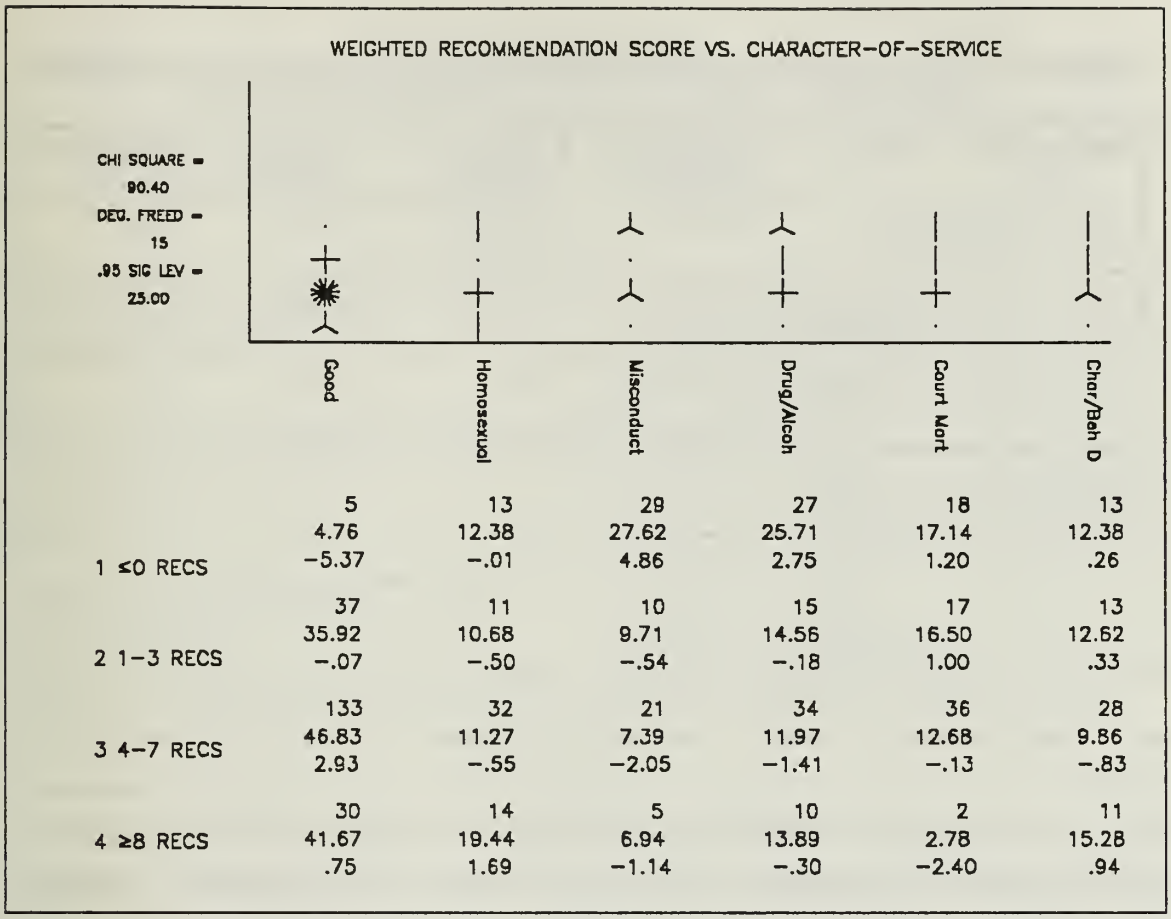
ADJUSTED RECOMMENDATION SCORE VS. CHARACTER-OF-SERVICE

CHI SQUARE = 84.55
DEG. FREED = 15
.95 SIG LEV = 25.00



	Good	Homosexual	Misconduct	Drug/Alcoh	Court Mart	Char/Beh D
1 ≤0 RECS	2	10	22	24	11	13
	2.44	12.20	26.83	29.27	13.41	15.85
	−5.09	−.06	4.08	3.25	.12	1.15
2 1−3 RECS	36	12	15	17	20	10
	32.73	10.91	13.64	15.45	18.18	9.09
	−.63	−.45	.65	.06	1.53	−.75
3 4−7 RECS	134	33	23	35	40	31
	45.27	11.15	7.77	11.82	13.51	10.47
	2.55	−.62	−1.90	−1.51	.27	−.53
4 ≥8 RECS	33	15	5	10	2	11
	43.42	19.74	6.58	13.16	2.63	14.47
	1.02	1.81	−1.27	−.47	−2.50	.76

Figure 15. The Cross-Tabulation of Adjusted Recs. vs. Character-of-Service (Case 3).



	Good	Homosexual	Misconduct	Drug/Alcoh	Court Mart	Char/Beh D
1 ≤0 RECS	5	13	28	27	18	13
	4.76	12.38	27.62	25.71	17.14	12.38
	−5.37	−.01	4.86	2.75	1.20	.26
2 1−3 RECS	37	11	10	15	17	13
	35.92	10.68	9.71	14.56	16.50	12.62
	−.07	−.50	−.54	−.18	1.00	.33
3 4−7 RECS	133	32	21	34	36	28
	46.83	11.27	7.39	11.97	12.68	9.86
	2.93	−.55	−2.05	−1.41	−.13	−.83
4 ≥8 RECS	30	14	5	10	2	11
	41.67	19.44	6.94	13.89	2.78	15.28
	.75	1.69	−1.14	−.30	−2.40	.94

Figure 16. The Cross-Tabulation of Wtd. Rec. Score vs. Character-of-Service (Case 4).

Table 10. **RESULTS OF CROSS-TABULATIONS INVOLVING RECOMMENDATIONS:** for figures 13 through 16.

Independent Variable	Chi-Square Statistic	Degrees of Freedom	.95 Level of Significance	Dependency?
Total Recommended (Case 1), for Figure 13	82.89	15	25.00	yes
Total Not Recommended (Case 2), for Figure 14	17.55	5	11.07	yes
Recommendation Score (Case 3), for Figure 15	84.55	15	25.00	yes
Weighted Recommendation Score (Case 4), for Figure 16	90.40	15	25.00	yes

It is interesting to note that the inclusion of the negative recommendation data does not strongly change the result of the test (see Figure 13, Figure 14, Figure 15, and Figure 16), although the chi-square test statistic does increase as negative recommendations are given greater weights. It seems logical that the negative recommendations should be included, because they affected approximately 15% of the scores. Including them is also more encompassing and should give a truer overall characterization.

Analysis of the residuals produced by the recommendation score cross-tabulations with their bootstrap residual values showed several significant relationships listed in Table 11 .

Table 11. RESIDUAL ANALYSIS OF RECOMMENDATION VS. CHARACTER-OF-SERVICE

Independent Variable	0.95 Quantile Bootstrap Residual Value	Significant Categorical Relationship	Residual Value
Number of Recommendations (Case 1)	3.18	No Recommendations - Good	-4.54
		No Recommendations - Misconduct	4.43
Recommendation Score (Case 3)	2.92	Score ≤ 0 - Good	-5.09
		Score ≤ 0 - Misconduct	4.08
		Score ≤ 0 - Drug Alcohol	3.25
Weighted Recommedation Score (Case 4)	3.04	Score ≤ 0 - Good	-5.37
		Score ≤ 0 - Misconduct	4.86

The residual analysis shows clearly that having at least one recommendation (case 1) is highly associated with the good character-of-service category (of 205 persons, only one had no recommendations at all). Similarly, when negative recommendations are added in (cases 3 and 4), a low number of negative recommendations is highly associated with this good category (only 15 out of 205 had any not recommended). This is shown by the residual analysis to be strongly significant.

Low recommendation scores were strongly associated with the misconduct and drug/alcohol abuse character-of-service categories. Low recommendation scores were not similarly associated with the other three adverse character-of-service categories (homosexual, court martial, or behavior disorder).

Note that in the total of recommendations (Figure 13), the homosexual character-of-service has a very large residual (3.09) associated with large numbers of recommendations. This is only slightly lower than the 95th quantile bootstrap residual value of 3.18, and as the third largest residual in the table, is probably significant. A large number of recommendations (8 or more) appears to be associated with this character-of-service. The high number of recommendations that those discharged with this character-of-service receive is very different from all other adverse categories.

F. FURTHER TESTING

The results shown above certainly indicate that a relationship exists between the amount and type of derogatory information and recommendations contained in the investigation records and the subsequent character of service category. More detailed analysis of other types of derogatory information is called for and was not accomplished because of time constraints.

IV. FURTHER ANALYSIS

A. GENERAL

The data analysis conducted to this point of the thesis dealt with the cross-tablation of categorical data, the chi-square test of differences of probabilities, and analysis of the residuals. There are many other ways to look at this data, some of which was briefly investigated and which warrant further investigation.

B. ANALYSIS OF PRODUCTIVITY OF SOURCES

Security background investigations are expensive operations. They involve many man-hours of investigators' time and much of the government's money to conduct. The efficient conduct of investigations is desireable not only to save money, but to avoid a large backload of investigations and to quickly award or reject the clearances; persons awaiting the clearance procedure cannot perform the job to which they are assigned until the clearance is granted. Is the quality and quantity of information obtained from all sources equal? Is it reasonable to spend equal amounts with each source of information?

There are several ways available to provide insight toward the answer to these questions. One method is to look at the total amount of information available from each source and to compare the amounts. Using APL logical operations, a total amount of information was obtained in three different categories:

1. The amount of derogatory information obtained from each source totalled across all records.
2. The number of recommendations obtained from each source totalled across all records;
3. The number of "not recommended" obtained from each source category totalled across all records.

These totals were graphically displayed using Grafstat's bar chart capability and the results are listed in Figure 17.

The factor which was immediately apparent to me was that neighbors appear to be much more reticent about discussing negative factors during a background investigation than persons from other source categories; or, are much more likely to say nice things about their neighbors and friends than are persons associated with the other source categories. In other words, based upon the bar charts, it may not be effective to

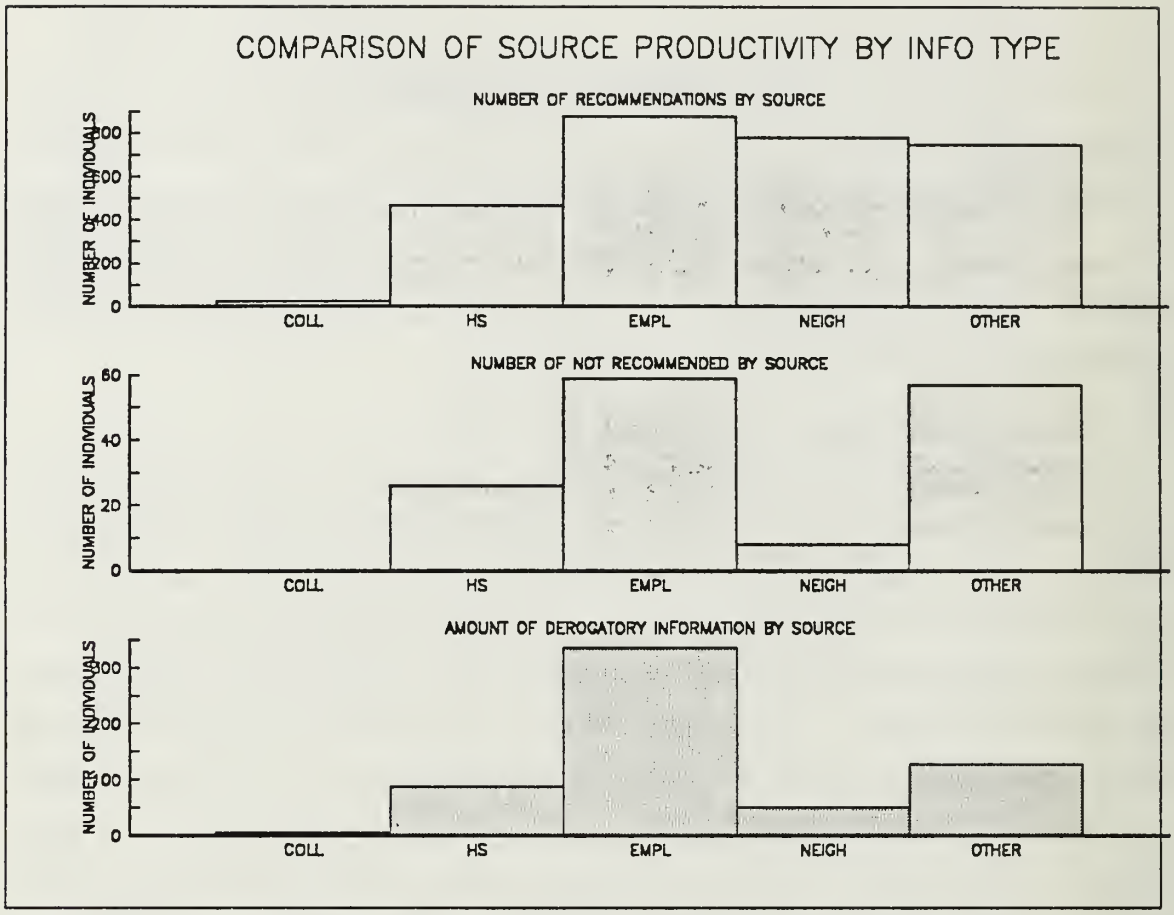


Figure 17. Productivity of Source Categories Across the Entire Data Base.

investigate subjects' neighbors in depth unless other areas of the investigation suggest that more detailed information might be obtained from neighbors.

Another factor apparent from the investigation is that little information is obtained from the college source category and that it may not be cost effective to expend resources at the college of a subject unless other portions of the investigation suggest that information can be obtained there.

Some important facts must be considered here before drawing specific conclusions from this figure. The data does not represent a random sample from the population at large, but a random sample from each discharge category according to an arbitrarily determined proportion. Each discharge category should be investigated separately to see if the trend apparent through the data base applies to all discharge categories in the same way.

A much more detailed mathematical investigation of this trend is needed to confirm the validity of this approach. The facts discussed above are strongly suggested by the figure, however, further analysis is needed in order to confirm or draw conclusions from them.

C. ANALYSIS OF THE WEIGHTED RECOMMENDATION SCORE

Chapter III, part E, discussed several techniques for evaluating the recommendation data disclosed as part of the investigation. Several scoring systems were discussed and detailed as cases 1, 3, and 4. Case 1 gave a direct score, with each recommendation worth one point, and all negative recommendations ignored. Case 3 was similar to case 1 but included all negative recommendations (recommendation with supervision, not recommended, and decline comment) scored as -1 and added to the score of case 1. Case 4, similarly awarded a score of + 1 for all positive recommendations, a score of -1 for the two "weak" negative recommendations (decline comment and recommend with supervision) and a -3 for strong negative recommendations.

Case 4 appears to be the better method of scoring for several reasons. It awards a higher penalty for the negative recommendation. This higher penalty appears to be justified when the relative rarity of negative recommendations is considered. Discussion with analysts from PERSEREC confirmed this approach as valid, given their experience. The highest chi-square statistic was also obtained in the test of this scoring method against the character of service.

The scoring technique provided a transformation from the strict categorical structure of the original data into a numerical range of values. This allows us to investigate the possible distribution of these scores.

The distribution of scores across all of the discharge categories together does not provide any particular insight into the meaning of the scores. However, the distribution of scores within each distribution category might show us some particular differences.

Grafstat, the mainframe computer statistical graphics package, contains a three-dimensional empirical density plot capability. The separate empirical density of the recommendation scores of each character-of-service category is displayed in a three-dimensional manner. This allowed rough comparisons between each empirical density to be made. Figure 18 displays the three-dimensional empirical density for the recommendation scores, with each "slice" representing a character of service category.

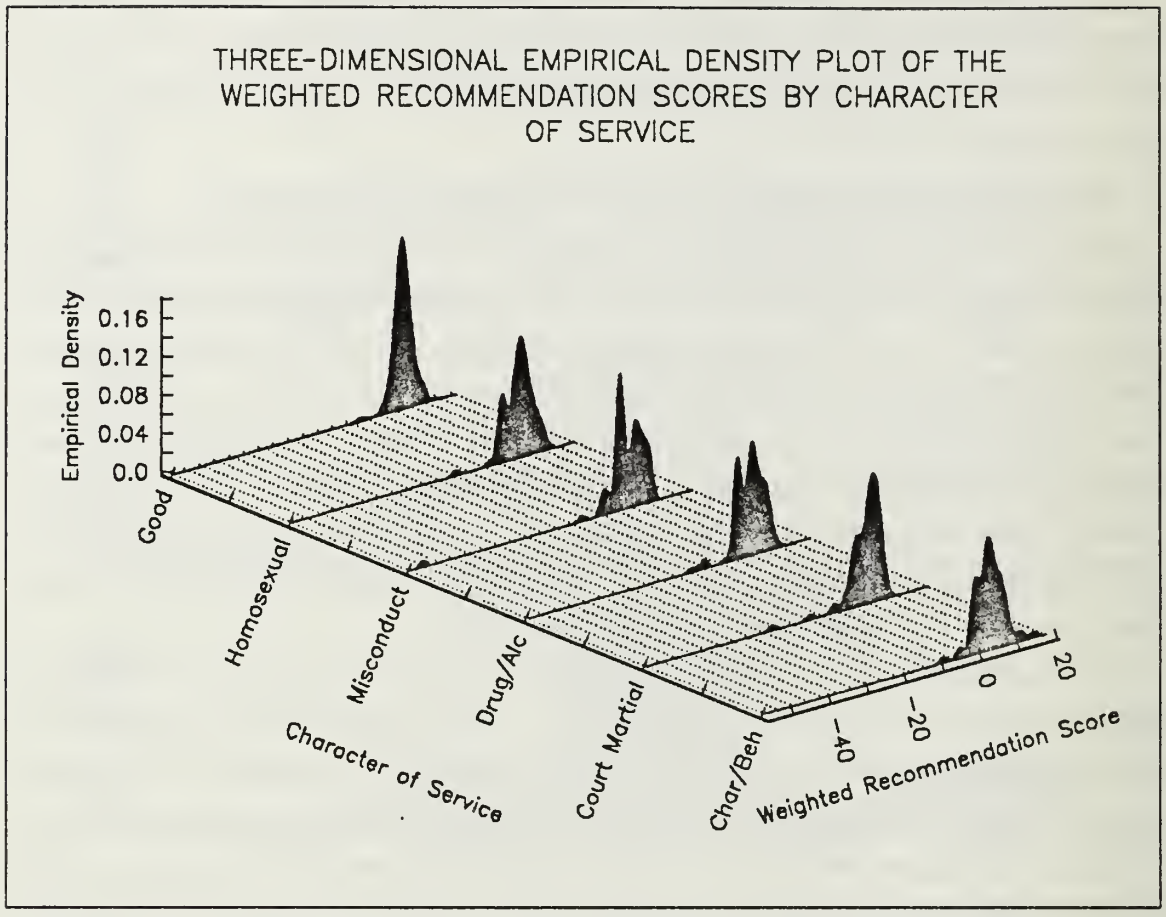


Figure 18. 3-D Empirical Density of the Weighted Recommendation Scores

The empirical density is a relatively simple idea. Each value within a sample (in this case the values are the recommendation scores) is thought to possess a density of $\frac{1}{n}$ where n represents the number of samples. Therefore the total density of the each empirical density plot represents $\frac{n}{n} = 1.0$. The density of each sample is distributed around the score according to a particular smoothing function. The smoothing function can be a uniform distribution (known as the boxcar function) or according to some other scheme (a smoothing function based upon the cosine function is commonly used). The density of each sample is calculated, and then the density at each location along the x-axis is calculated. One could say that the density from each sample is spread around each point for some distance. The total density at any particular point is the sum of any densities which overlap at that point. The empirical density plot allows us to get a rough

idea of how the sample is grouped about particular values and may assist in determining the distribution of the sample.

The three-dimensional empirical density plot of the recommendation scores also allows us to do something very practical in relation to the recommendation score. Figure 19 is the two-dimensional representation of the three-dimensional empirical density plot. Each 'slice' is laid flat in the two-dimensional plane. This representation allows us to interpret the x-axis values more precisely. What was immediately apparent was that the recommendation scores associated with the good character of service were closely grouped about a score of +6. There are very few negative values at all. The densities of the adverse character of service categories are much more spread out and have significantly more density in the negative score range. Note particularly the number of enlisted men who had very negative weighted recommendation scores and who were subsequently discharged for misconduct.

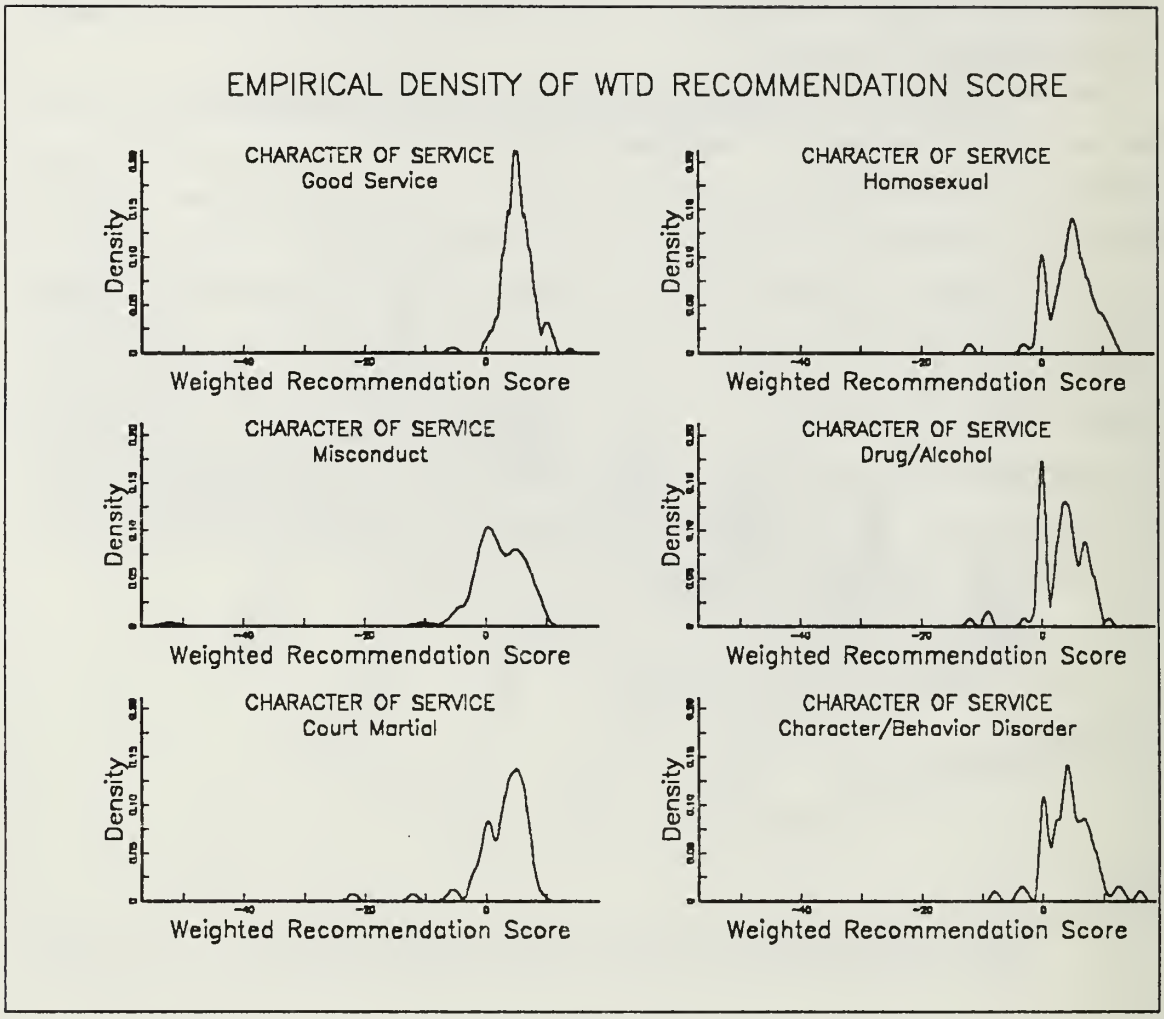


Figure 19. Empirical Density Plot of Weighted Rec. Score by Category

A quick visual analysis of the empirical density plots suggests that the recommendation scoring technique could assist the adjudicator in evaluating a investigation record. If all persons with a recommendation score of less than zero were denied clearance, we could be reasonably assured that we would reject few persons who would complete their term of service successfully; however, we could also be reasonably assured to reject a significant number of people who would later be discharged for adverse reasons. Of course, this consideration must be weighted against the high proportion of the population who receive the good character-of-service versus the relatively low number who receive adverse characters-of-service. Two percent of the good category probably consists of many more people than 44% of the misconduct category. Table 12 displays the

percentage of each character-of-service category who had a non-positive weighted recommendation score. The percentage of the general population who receive each of the six character-of-service discharge categories is shown for comparison.[1]

Table 12. PERCENTAGE OF NON-POSITIVE RECOMMENDATION SCORES

Character-of-Service Discharge Category	Number in Sample.	Number with *Score* ≤ 0	Percentage with *Score* ≤ 0	Percentage Who Receive This Category in General Population
Good	205	5	2.43%	90.4%
Homosexual	70	13	18.57%	0.92%
Misconduct	65	29	44.61%	1.2%
Drug Alcohol Abuse	86	27	31.39%	1.8%
Court Martial	73	18	24.65%	0.18%
Character and Behavior Disorder	65	13	20.00%	0.65%

Further analysis of the recommendation scoring methods is clearly indicated. Recommendation data makes up a large proportion of this data base. Comparisons between recommendations and derogatory information may also yield important information. It is important to note that the vast majority of persons who are investigated for a security clearance finish their service successfully and the adverse discharges are not common. The proportion of records in each character-of-service category is not in any relation to their appearance in the actual population.

---

1 For all consideration in this thesis, the "general population" consists of those who have security investigations conducted.

V. CONCLUSIONS AND RECOMMENDATIONS

A. FINDINGS

1. Cross Tabulation and Residual Analysis

a. General Findings

There is clear indication that the distribution of investigative data is different for each character-of-service category, especially for each adverse category when compared to the good character-of-service.

The good character-of-service category is generally described by sparse derogatory information, very few negative recommendations and a relatively high recommendation score. In fact, persons receiving this characterization of service usually had several or more recommendations.

The misconduct, drug/alcohol abuse and court-martial character of service category generally exhibit an opposite trend in regard to derogatory information and recommendations. These categories tend to have significantly more derogatory information and significantly lower recommendation scores.

The homosexual and character and behavior disorder categories exhibit the general trend of the other adverse character-of-service categories, however, their values tend to be much less pronounced. It is more difficult to generalize their attributes with those of the other categories.

b. Specific Findings

Age of the investigation subject bears no particular relation to any character-of-service category.

Female service-members comprise a significantly lower percentage of the drug/acohol abuse and court-martial character-of-service categories; however, women display a significantly higher percentage of the character and behavior disorder category.

The lack of a high school diploma is significantly associated with the court martial character-of-service, but is not significantly associated with any of the other categories.

The lack of any derogatory information is significantly associated with only the good character-of-service category, however, 40% of persons in this category had at least one item of derogatory information in their file. The number of people with at

least one item of derogatory information in the adverse categories tended to be around 80%.

Drug abuse disclosed in the investigation is significantly associated with the drug/alcohol character-of-service category. Nearly 50% of those whose service was terminated because of drug abuse had evidence of drug abuse disclosed in their investigations. The misconduct category showed similar, and only slightly less significant, association. Persons who completed service with a good character-of-service evidence significantly lower evidence of drug use in their investigations. It is apparent that drug abuse disclosed in the investigation should be carefully weighed and could prove to be a major discriminator in the adjudication of a clearance.

The proportion of recommendations varied significantly among all the character-of-service categories. In fact, a remarkable attribute of those who received a good character-of-service is that only one out of 205 in the sample received no recommendations at all.

The recommendation scoring technique appears to be a valuable method of generalizing the recommendations received. This scoring system is better than looking at positive or negative recommendations alone because it involves more of the data and provides a means of balancing between the two extremes. The cross-tabulation of the recommendation scoring showed that a low score is not a characteristic of the good character-of-service.

2. Productivity of Sources

A brief look at the amount of derogatory information, number of recommendations, and number of negative recommendations shows that the neighborhood and college sources may not be the best place to expend limited investigative resources compared with high schools, employers, and the other developed sources.

3. Weighted Recommendation Score as a Predictor

The brief look at the three-dimensional empirical density plot of the weighted recommendation scores indicates that this scoring method may prove valuable when adjudicating investigations. Based upon this data, rejecting any subject with a score less than 0 would have a negligible effect upon the persons who later receive a good character-of-service while eliminating a significant number of persons who would later have received an adverse discharge.

B. RECOMMENDATIONS FOR FURTHER STUDY INVOLVING THIS DATA

The actual data analysis conducted in this thesis was of those factors which appeared obvious. There is much more analysis which can and should be done. The data is now configured for ease of use by someone with an elementary knowlege of APL and Grafstat.

The recommendations are:

1. Continued research should be conducted into the relationships between the character-of-service categories and particular groupings of investigation data.
2. Further investigation is warranted into various productivity of source issues hinted at during this investigation. Investigations into productivity of source by information type, and how the sources relate to the various character-of-service categories would appear to be fertile areas available for analysis.
3. The recommendation scoring techniques should be further analyzed to find an optimal score value for each type of recommendation. It seems reasonable that this scoring system could be very useful as another tool available to the investigation adjudicator when evaluating investigation records.

C. RECOMMENDATIONS FOR FURTHER STUDIES

The work accomplished in this thesis has shown that considerable manipulation of a large data base can be accomplished using APL. With proper design, and the newer versions of APL (APL2) and faster mainframes and microcomputers even larger data sets can be efficiently manipulated and analyzed.

A study involving a large sample from the investigative population at large should be conducted. Investigations similar to what has been conducted here along with analysis of attributes of the general population should be conducted.

APPENDIX A. FREQUENCY TABULATIONS OF EACH VARIABLE

This appendix contains frequency tabulations of each categorical (numerically coded) variable. Some variables are given alpha-numeric names (for example AA1), because that particular source provided multiple pieces of information for some records. These tabulations were performed using the Codebook Procedure of Statgraphics (version 2.6). The tabulations headings were edited to make them easier to understand. Labels are limited to 10 characters within Statgraphics.

Variable C, Sex

Sex	Frequency	Codes
Male	479	1
Female	85	2

Variable I, Reason for Background Investigation

Reason	Frequency	Codes
Acc Class	315	1
Nuc Wpn	112	2
SCI	5	3
Pres Spt	132	4

Variable J, Navy Occupation Code

Occupation	Frequency	Codes
BE/E	6	1
CTA	9	11
CTI	17	12
CTM	13	13
CTO	15	14
CTR	15	15
CTR/T	7	16
CTT	28	17
CTT/R	1	18
GM	1	21
GMT	38	22
IS	21	31
MS	2	41
PE	33	51
PS	4	52

RM	269	61
SW	1	71
SWSE	11	72
TM	4	81
TMS	34	82
TMT	18	83
UFT	17	91

Variable K, Reason for Interview

Reason	Frequency	Codes
Int B Info (IBI)	79	1
Stmt Ps Hs (SPH)	15	2
Not Conducted	276	3
Required	118	4
Unfav. Inf	76	5

Variable L1, Interview Information #1

Information	Frequency	Codes
Brk & Ent	1	1000
Burglary	2	1001
Robbery	1	1042
Theft	6	1043
Sus Theft	2	1044
Assault	3	1050
Rape	1	1052
Alcohol Pr	4	1100
DUI 1	2	1101
Liq Law Vi	1	1104
Min Alc Po	2	1105
PubIntox>1	1	1107
Marij Arr	3	1120
Marij 1-5	72	1121
Marij 6-25	6	1122
Marij >25	2	1123
Susp Drug	2	1140
DrgUse 1-5	7	1141
Credit	4	1210
Bad Checks	2	1230
Unpd Bills	15	1233
Homo Susp	2	1422
Tresspass	1	1539
Fired	15	1561
Suspnd Sch	7	1564
SusCommSym	2	1600
Depression	1	1723
Leave Acct	2	1804
Mov Veh 1	8	1810
Mov Veh2-5	5	1811

Mov Veh >5	1	1812
Dmg Per Pr	2	1841
Juven Rec	6	1843
Runaway	2	1845
Voyeur	1	1847
Job NotSPH	2	1923
No Informa	368	9999

Variable L2, Interview Information #2

Information	Frequency	Codes
Brk & Ent	2	1000
Burglary	1	1001
Theft	3	1043
Assault	2	1050
Battery	1	1051
Alcohol Pr	11	1100
DUI 1	1	1101
DUI >1	1	1102
Min Alc Po	2	1105
PubIntox 1	1	1106
Marij Arr	4	1120
Marij 1-5	14	1121
Marij 6-25	1	1122
Susp Drug	1	1140
DrgUse 1-5	4	1141
Sale Drugs	3	1150
Drug Waivr	2	1160
Credit	2	1210
Bad Checks	3	1230
Sued Nonpa	1	1232
Unpd Bills	5	1233
Veh Reposs	1	1235
Homo Admit	1	1421
Disrespect	1	1530
Disag Empl	2	1532
Perf Unsat	1	1538
Disenr Sch	1	1540
Fired	10	1561
Suspnd Sch	6	1564
Absnt Freq	3	1570
Immat Beha	1	1573
Emot Prob	1	1701
Depression	1	1723
Leave Acct	1	1804
Mov Veh 1	5	1810
Mov Veh2-5	1	1811
Stnd Veh 1	2	1813
Disord Con	3	1821
Dstb Peace	1	1822
Harrassmnt	1	1823
Lisc Suspd	3	1825
Crim Misch	2	1840

Juven Rec	5	1843
Mal Mischf	1	1844
Runaway	1	1845
Fraud Enli	1	1922
Job NotSPH	3	1923

Variable L3, Interview Information #3

Information	Frequency	Codes
Robbery	6	1043
Sus Theft	1	1044
Assault	2	1050
Incorrigbl	1	1070
Alcohol Pr	6	1100
DUI 1	1	1101
Drnk & Dis	1	1103
Liq Law Vi	1	1104
Min Alc Po	3	1105
Marij Arr	3	1120
Marij 1-5	5	1121
DrgUse 1-5	1	1141
Unpd Bills	2	1233
Juv Homo	1	1423
Perf Unsat	1	1538
Fired	1	1561
Absnt Freq	3	1570
Tardy	1	1577
SusCommSym	1	1600
Mov Veh 1	6	1810
Dstb Peace	1	1822
Lisc Suspd	1	1825
Vandalism	1	1829
Crim Misch	1	1840
Dmg Per Pr	1	1841
Juven Rec	3	1843
Mal Mischf	1	1844
Add NotSPH	1	1920
Job NotSPH	1	1923

Variable L4, Interview Information

Information	Frequency	Codes
Robbery	2	1043
Lied	1	1071
Alcohol Pr	2	1100
DUI 1	1	1101
Min Alc Po	1	1105
Pub Intox1	1	1106
Marij 1-5	2	1121
Susp Drug	1	1140

DrgUse 1-5	3	1141
Drug Waivr	1	1160
Homo Susp	1	1422
Fem Impers	1	1430
Un Absense	1	1501
Fired	1	1561
Emot Prob	1	1701
Mov Veh2-5	1	1811
Lisc Suspd	1	1825
Juven Rec	1	1843
Mal Mischf	1	1844

Variable M1, NAC Checks Findings #1

Information	Frequency	Codes
Brk & Ent	1	1000
Burglary	2	1001
Forgery	1	1010
Theft	4	1043
Assault	1	1050
Rape	1	1052
DUI 1	2	1101
Marij Arr	2	1120
Marij 1-5	1	1121
DrgUse 1-5	1	1141
Bad Checks	2	1230
Homo Susp	1	1422
Vio Parole	1	1502
Not El Reh	1	1537
Tng Dischg	1	1543
Mov Veh 1	1	1810
Disord Con	1	1821
Harrassmnt	1	1823
Dmg Per Pr	1	1841
No Informa	538	9999

Variable M2, National Agency Check Findings

Information	Frequency	Codes
Burglary	4	1001
Robbery	1	1042
Theft	1	1043
DUI 1	1	1101
Tresspass	1	1539
Reck Drive	1	1805
Disord Con	1	1821
Juven Rec	1	1843

Variable N1, Local Agency Check Findings #1

Information	Frequency	Codes
Burglary	6	1001
Larceny	1	1040
Theft	11	1043
Sus Theft	1	1044
Assault	3	1050
Rape	1	1052
Alcohol Pr	1	1100
DUI 1	6	1101
DUI>1	2	1102
Liq Law Vi	5	1104
Min Alc Po	10	1105
PubIntox 1	3	1106
PubIntox>1	1	1107
Marij Arr	4	1120
Marij 1-5	2	1121
Marij 6-25	1	1122
DrgUse 1-5	2	1141
Bad Checks	5	1230
Sued Nonpa	2	1232
Unpd Bills	1	1233
Tresspass	2	1539
Hit & Run	1	1803
Leave Acct	3	1804
Reck Drive	3	1805
Mov Veh 1	45	1810
Mov Veh2-5	29	1811
Mov Veh >5	3	1812
Stnd Veh 1	1	1813
Disord Con	2	1821
Crim Misch	1	1840
Dmg Per Pr	2	1841
Juven Rec	31	1843
No Informa	373	9999

Variable N2, Local Agency Check Findings #2

Information	Frequency	Codes
Brk & Ent	5	1000
Burglary	1	1001
Theft	9	1043
Battery	3	1051
Incorrigbl	1	1070
Alcohol Pr	2	1100
DUI 1	4	1101
DUI>1	1	1102
Liq Law Vi	1	1104
Min Alc Po	1	1105
PubIntox 1	2	1106
Marij Arr	1	1120

Marij 1-5	3	1121
Marij 6-25	1	1122
DrgUse 1-5	1	1141
Bad Checks	1	1230
No Restit	1	1231
Sued Nonpa	1	1232
Homo Susp	1	1422
Vio Parole	1	1502
Fail Aper	10	1572
Leave Acct	1	1804
Reck Drive	3	1805
Mov Veh 1	8	1810
Mov Veh2-5	7	1811
Stnd Veh 1	1	1813
Disord Con	2	1821
Dstb Peace	1	1822
Harrassmnt	1	1823
Il Firearm	1	1824
Lisc Suspd	8	1825
Out Warrnt	2	1828
Crim Misch	2	1840
Dmg Per Pr	1	1841
Juven Rec	6	1843
Mal Mischf	1	1844
Voyeur	1	1847

Variable N3, Local Agency Check Findings #3

Information	Frequency	Codes
Burglary	1	1001
Larceny	2	1040
Theft	4	1043
Assault	3	1050
Alcohol Pr	1	1100
DUI 1	2	1101
Liq Law Vi	1	1104
PubIntox 1	2	1106
DrgUse 1-5	1	1141
Sale Drugs	1	1150
Fem Impers	1	1430
Tresspass	2	1539
Fail Aper	1	1572
Mov Veh 1	6	1810
Mov Veh2-5	3	1811
Mov Veh >5	1	1812
Disord Con	3	1821
Il Firearm	1	1824
Lisc Suspd	3	1825
Juven Rec	1	1843

Variable N4, Local Agency Check Findings #4

Information	Frequency	Codes
Burglary	1	1001
Larceny	1	1040
R StolProp	1	1041
Theft	1	1043
Drnk & Dis	1	1103
Liq Law Vi	1	1104
Marij Arr	1	1120
Susp Drug	1	1140
Unpd Bills	1	1233
Fail Aper	1	1572
Leave Acct	1	1804
Mov Veh2-5	2	1811
Harrassmnt	1	1823
Juven Rec	1	1843

Variable O1, Credit Bureau Check Findings #1

Information	Frequency	Codes
Bad Checks	3	1230
Unpd Bills	22	1233
Veh Reposs	1	1235
No Informa	538	9999

Variable O2, Credit Bureau Check Findings #2

Information	Frequency	Codes
Bad Checks	1	1230
No Restit	1	1231
Unpd Bills	2	1233
Bankruptcy	1	1250

Variable P, High School # of Sources

# of Sources	Frequency	Codes
0	290	0
1	111	1
2	91	2
3	42	3
4	21	4
5	6	5
6	2	6
7	1	7

Variable Q1, High School Recommendations

Type Recomm.	Frequency	Codes
1 Recomm	116	11
2 Recomm	85	12
3-5 Recomm	59	13
6-10Recomm	1	14
1 RecWSu	2	21
1 DecCom	1	31
3-5 DecCom	1	33
1 NotRec	7	41
2 NotRec	2	42
No Intervw	290	99

Variable Q2, H.S. Recommendation #2

Type Recomm.	Frequency	Codes
1 Recomm	3	21
1 RecWSu	10	31
2 RecWSu	1	32
1 NotRec	6	41
2 NotRec	2	42

Variable Q3, HS Recommendations #3

Type Recomm.	Frequency	Codes
1 RecWSu	2	21
1 DecCom	3	31
2 DecCom	1	41
2 NotRec	2	42

Variable R1, HS Diploma Status

Diploma Status	Frequency	Codes
Dipl Grad	454	1
Non Grad	81	2
Gen Eq Dip	29	3

Variable R2, High School Findings #1

Information	Frequency	Codes
Theft	1	1043

Assault	1	1050
Incorrigbl	1	1070
Alcohol Pr	3	1100
DUI >1	1	1102
Marij 1-5	3	1121
Susp Drug	2	1140
DrgUse 1-5	1	1141
Insubordin	2	1535
Perf Unsat	1	1538
Lack Motiv	1	1563
Suspnd Sch	3	1564
Absnt Freq	12	1570
Braggart	1	1571
Immat Beha	10	1573
Needs Supv	1	1575
Emot Stab?	1	1702
Clmd Diplm	3	1921
No Informa	6	9999

Variable R3, High School Findings #2

Information	Frequency	Codes
Theft	1	1043
Dishonest	2	1060
Marij 1-5	1	1121
Susp Drug	2	1140
DrgUse 1-5	2	1141
Disrespect	1	1530
Insubordin	1	1535
Tng Dischg	1	1543
Lack Motiv	3	1563
Suspnd Sch	5	1564
Absnt Freq	2	1570
Braggart	1	1571
Tardy	3	1577
Undependbl	1	1578
Emot Stab?	1	1702
Vandalism	1	1829

Variable R4, High School Findings #3

Information	Frequency	Codes
Theft	1	1043
Alcohol Pr	1	1100
Susp Drug	2	1140
Disrespect	1	1530
Expelled	1	1560
Absnt Freq	1	1570
Undependbl	1	1578
Emot Stab?	2	1702

Clmd Diplm	1	1921

Variable S, College # of Sources

# of Sources	Frequency	Codes
0	547	0
1	9	1
2	5	2
3	2	3
4	1	4

Variable T, College Recommendations

Type Recomm.	Frequency	Codes
1 Recomm	8	11
2 Recomm	5	12
3-5 Recomm	2	13
1 DecCom	1	31
No Intervw	548	99

Variable U, College Findings

Information	Frequency	Codes
Assault	1	1050
Lied	1	1071
Unpd Bills	2	1233
Empl False	1	1924
No Intervw	559	9999

Variable V, Employer # of Sources

# of Sources	Frequency	Codes
0	191	0
1	132	1
2	109	2
3	75	3
4	34	4
5	10	5
6	6	6
7	4	7
8	2	8
11	1	11

Variable W, Co-worker # of Sources

# of Sources	Frequency	Codes
0	356	0
1	118	1
2	58	2
3	24	3
4	4	4
5	3	5
6	1	6

Variable X1, Co-worker Recommendations #1

Type Recomm.	Frequency	Codes
1 Recomm	89	11
2 Recomm	91	12
3-5 Recomm	162	13
6-10Recomm	16	14
11+ Recomm	1	15
2 RecWSu	1	22
1 DecCom	12	31
2 DecCom	6	32
3-5 DecCom	2	33
1 NotRec	8	41
2 NotRec	1	42
3-5 NotRec	1	43
No Intervw	174	99

Variable X2, Employment Recommendations #2

Type Recomm.	Frequency	Codes
1 Recomm	1	11
2 Recomm	1	12
3-5 Recomm	3	13
1 RecWSu	1	21
2 RecWSu	1	22
1 DecCom	42	31
2 DecCom	15	32
3-5 DecCom	3	33
1 NotRec	13	41
2 NotRec	4	42
3-5 NotRec	4	43
6-10NotRec	1	44

Variable X3, Employment Recommendations #3

Type Recomm.	Frequency	Codes
1 DecCom	6	31
6-10DecCom	1	34
1 NotRec	3	41
2 NotRec	2	42

Variable Y1, Employment Findings #1

Information	Maximum Frequency	Codes
Theft	1	1043
Sus Theft	3	1044
Lied	2	1071
Alcohol Pr	2	1100
Marij Arr	1	1120
Marij 1-5	1	1121
Susp Drugs	2	1140
DrgUse 1-5	1	1141
Bad Checks	1	1230
Disrespect	1	1530
Not El Reh	18	1537
Perf Unsat	3	1538
Fired	77	1561
Freq Quit	7	1562
Absnt Freq	7	1570
Immat Beha	8	1573
Lack Judge	1	1574
Needs Supv	1	1575
Tardy	1	1577
Undependbl	5	1578
Juven Rec	1	1843
Job NotSPH	5	1923
Time Unact	20	1924
Empl False	3	1927
No Intervw	392	9999

Variable Y2, Employment Findings #2

Information	Frequency	Codes
Theft	5	1043
Sus Theft	1	1044
Assault	1	1050
Alcohol Pr	2	1100
Marij 1-5	5	1121
DrgUse 1-5	1	1141
Unpd Bills	1	1233
Disrespect	1	1530
Disag Em	3	1532
Emplt Prob	1	1533
Insubordin	1	1535

Misconduct	1	1536
Not El Reh	8	1537
Perf Unsat	11	1538
Fired	7	1561
Freq Quit	1	1562
Lack Motiv	5	1563
Absnt Freq	33	1570
Immat Beha	6	1573
Tardy	1	1577
Undependbl	4	1578
Emot Prob	1	1701
Emot Stab?	1	1702
Empl False	3	1924

Variable Y3, Employment Findings #3

Information	Frequency	Codes
Forgery	1	1010
Theft	3	1043
Sus Theft	2	1044
Alcohol Pr	4	1100
Marij 1-5	1	1121
Susp Drugs	1	1140
Bad Checks	1	1230
Homo Susp	1	1422
Disrespect	1	1530
Insubordin	1	1535
Not El Reh	6	1537
Perf Unsat	1	1538
Tresspass	2	1539
Fired	1	1561
Lack Motiv	1	1563
Absnt Freq	5	1570
Braggart	1	1571
Immat Beha	3	1573
Tardy	5	1577
Undependbl	2	1578
Emot Stab?	1	1702
Nerv Condn	1	1712
Empl False	3	1924

Variable Y4, Employment Findings #4

Information	Frequency	Codes
Lied	1	1071
Marij 1-5	2	1121
Susp Drugs	1	1140
Homo Susp	1	1422
Insubordin	1	1535
Not El Reh	1	1537

Immat Beha	2	1573
Undependbl	1	1578
Com Threat	1	1820

Variable Z, Neighborhood # of Sources

# of Sources	Frequency	Codes
0	173	0
1	113	1
2	140	2
3	66	3
4	39	4
5	23	5
6	9	6
7	1	7

Variable AA1, SPH Listed Neighborhood Source Recommendations

Type Recomm.	Frequency	Codes
1 Recomm	83	11
2 Recomm	81	12
3-5 Recomm	76	13
6-10Recomm	3	14
1 DecCom	1	31
1 NotRec	2	41
No Intervw	318	99

Variable AA2, Developed Neighborhood Recommendations #1

Type Recomm.	Frequency	Codes
1 Recomm	38	11
2 Recomn	50	12
3-5 Recomm	46	13
6-10Recomm	2	14
1 RecWSu	1	21
1 DecCom	12	31
2 DecCom	8	32
3-5 DecCom	2	33
1 NotRec	5	41
None Devel	400	99

Variable AA3, Developed Neighborhood Recommendations #2

Type Recomm.	Frequency	Codes

1	DecCom	10	31
2	DecCom	2	32
3-5	DecCom	2	33
1	NotRec	1	41

Variable AB1, Neighborhood Findings #1

Information	Frequency	Codes
Theft	1	1043
Wild (Beh)	1	1091
Alcohol Pr	1	1100
DUI 1	1	1101
Marij 1-5	5	1121
Sus Theft	1	1140
DrgUse 1-5	1	1141
Unpd Bills	3	1233
Disrespect	1	1530
Perf Unsat	1	1538
Tresspass	2	1539
Lack Motiv	1	1563
Braggart	1	1571
Immat Beha	4	1573
Undependbl	2	1578
Suicide At	1	1704
Juven Rec	1	1843
Mal Mischf	1	1844
Voyeur	1	1847
No Intervw	534	9999

Variable AB2, Neighborhood Findings #2

Information	Frequency	Codes
Brk & Ent	1	1000
Theft	1	1043
Wild (Beh)	1	1091
Alcohol Pr	1	1100
Susp Drugs	5	1140
DrgUse 1-5	2	1141
Unpd Bills	2	1233
Attitud Pr	2	1530
Undependbl	1	1578
Depression	1	1723

Variable AB3, Neighborhood Findings #3

Information	Frequency	Codes
Burglary	1	1001

Marij 1-5	1	1121
Juven Rec	1	1843

Variable AC, Other # of Sources

# of Sources	Frequency	Codes
0	205	0
1	98	1
2	107	2
3	86	3
4	42	4
5	12	5
6	8	6
7	3	7
8	2	8
17	1	17

Variable AD1, Other Recommendations #1

Type Recomm.	Frequency	Codes
1 Recomm	95	11
2 Recomm	105	12
3-5 Recomm	126	13
6-10Recomm	10	14
1 RecWSu	3	21
1 DecCom	4	31
2 DecCom	1	32
1 NotRec	5	41
2 NotRec	3	42
3-5 NotRec	7	43
No Intervw	205	99

Variable AD2, Other Recommendations #2

Type Recomm.	Frequency	Codes
1 Recomm	2	11
2 Recomm	2	12
1 DecCom	5	31
3-5 DecCom	1	33
1 NotRec	10	41
2 NotRec	2	42
11+ NotRec	1	45

Variable AD3, Other Recommendations #3

Type Recomm	Frequency	Codes
3-5 DecCom	1	33

Variable AE1, Other Findings #1

Information	Frequency	Codes
Brk & Ent	1	1000
R StolProp	1	1041
Theft	1	1043
Sus Theft	2	1043
Lied	2	1071
Alcohol Pr	2	1100
Min Alc Po	1	1105
Marij Arr	1	1120
Marij 1-5	12	1121
Susp Drugs	3	1140
DrgUse 1-5	1	1141
Bad Checks	2	1230
Unpd Bills	2	1233
Disq Nuc	2	1561
Lack Motiv	1	1563
Absnt Freq	1	1570
Braggart	1	1571
Immat Beha	6	1573
Needs Supv	1	1575
Undependbl	4	1578
Emot Prob	1	1701
Emot Stab?	1	1702
Psych Prob	1	1703
Anxiety	1	1710
Juven Delq	1	1842
Juven Rec	4	1843
Time Unact	1	1927
No Intervw	507	9999

Variable AE2, Other Findings #2

Information	Frequency	Codes
R StolProp	1	1041
Theft	4	1043
Lied	4	1071
Wild (Beh)	1	1091
Alcohol Pr	6	1100
DUI 1	1	1101
Min Alc Po	1	1105
Maij 1-5	1	1121
Susp Drugs	1	1140
DrgUse 1-5	3	1141
Sale Drugs	1	1150

No Resti	1	1231
Homo Susp	1	1422
Disrespect	1	1530
Suspnd Sch	1	1564
Absnt Freq	2	1570
Braggart	1	1571
Immat Beha	1	1573
Easi Upset	3	1700
Emot Stab?	2	1702
Suicide At	1	1704
Depression	1	1723
Runaway	1	1845

Variable AE3, Other Findings #3

Information	Frequency	Codes
Theft	3	1043
Dishonest	1	1060
Violent	1	1090
Alcohol Pr	1	1100
Marij 1-5	2	1121
Susp Drugs	2	1140
Fired	1	1561
Immat Beha	3	1573
Reckl Cond	1	1576
Undependbl	3	1578
Emot Stab?	3	1702

Variable AE4, Other Findings #4

Information	Frequency	Codes
Forgery	2	1010
Theft	1	1043
Assault	1	1050
Battery	1	1051
Alcohol Pr	1	1100
Unpd Bills	1	1233
Lack Motiv	1	1563
Immat Beha	1	1573
Emot Prob	1	1701

Variable AF, Race

Race	Frequency	Codes
Caucasian	397	1
Black	50	2
Hispanic	11	3

Oriental	2	4
Unknown	104	5

Variable AG, Marital Status

Status	Frequency	Codes
Unmarried	517	1
Married	33	2
Divorced	14	4

Variable AJ, # of Dependents

# of Dependents	Frequency	Codes
0	521	0
1	30	1
2	11	2
3	2	3

Variable AN, # of Siblings

# of Siblings	Frequency	Codes
0 Siblings	24	0
1 Siblings	93	1
2 Siblings	118	2
3 Siblings	125	3
4 Siblings	82	4
5 Siblings	56	5
6 Siblings	14	6
7 Siblings	17	7
8 Siblings	9	8
9 Siblings	4	9
10 Sibling	4	10
12 Sibling	1	12
Unkn # Sib	17	50

Variable AO, Permanent Residence

State	Frequency	Codes
Alabama	18	1
Arizona	8	4
Arkansas	5	5
California	38	6
Colorado	6	8
Connecticu	9	9

DC	1	11
Florida	24	12
Georgia	11	13
Hawaii	2	15
Idaho	3	16
Illinois	20	17
Indiana	15	18
Iowa	7	19
Kansas	2	20
Kentucky	7	21
Louisiana	6	22
Maine	3	23
Maryland	10	24
Massachuse	20	25
Michigan	24	26
Minnesota	4	27
Mississipp	7	28
Missouri	14	29
Nebraska	3	31
Nevada	5	32
New Hampsh	2	33
New Jersey	19	34
New Mexico	5	35
New York	41	36
North Caro	15	37
Ohio	34	39
Oklahoma	9	40
Oregon	6	41
Pennsylvan	45	42
Puerto Ric	1	43
Rhode Isla	5	44
South Caro	15	45
South Dako	5	46
Tennessee	9	47
Texas	36	48
Virginia	19	51
Washington	6	53
West Virgi	8	54
Wisconsin	9	55
Wyoming	3	56

Variable AR, Age at enlistment

Age	Frequency	Codes
0	1	0
16.7	1	16.7
17	7	17.0
17.1	2	17.1
17.2	3	17.2
17.3	7	17.3
17.4	8	17.4
17.6	7	17.6
17.7	9	17.7

17.8	27	17.8
17.9	13	17.9
18	25	18.0
18.1	15	18.1
18.2	23	18.2
18.3	18	18.3
18.4	20	18.4
18.5	18	18.5
18.6	25	18.6
18.7	17	18.7
18.8	24	18.8
18.9	17	18.9
19	22	19.0
19.1	10	19.1
19.2	17	19.2
19.3	16	19.3
19.4	11	19.4
19.5	7	19.5
19.6	10	19.6
19.7	10	19.7
19.8	7	19.8
19.9	7	19.9
20	12	20.0
20.1	8	20.1
20.2	3	20.2
20.3	6	20.3
20.4	7	20.4
20.5	3	20.5
20.6	8	20.6
20.7	7	20.7
20.8	3	20.8
20.9	6	20.9
21	10	21.0
21.1	1	21.1
21.2	1	21.2
21.3	2	21.3
21.5	6	21.5
21.6	3	21.6
21.7	3	21.7
21.8	3	21.8
21.9	2	21.9
22	3	22.0
22.1	3	22.1
22.2	1	22.2
22.3	5	22.3
22.4	4	22.4
22.5	1	22.5
22.6	2	22.6
22.7	3	22.7
22.8	1	22.8
22.9	2	22.9
23	2	23.0
23.1	2	23.1
23.2	2	23.2
23.3	2	23.3
23.4	4	23.4

23.5	1	23.5
23.6	1	23.6
23.9	2	23.9
24.1	2	24.1
24.2	2	24.2
24.4	1	24.4
24.8	1	24.8
25	1	25.0
25.3	2	25.3
25.5	1	25.5
25.6	3	25.6
25.8	1	25.8
27	1	27.0
27.1	1	27.1
27.9	1	27.9
28	1	28.0
28.1	1	28.1
28.3	1	28.3
28.4	1	28.4
29.3	1	29.3
29.5	1	29.5
29.8	1	29.8
31.4	1	31.4

Variable AS, Months Between HS and Enlistment

Months	Frequency	Codes
0	48	0
1	52	1
2	40	2
3	30	3
4	22	4
5	10	5
6	20	6
7	22	7
8	25	8
9	14	9
10	16	10
11	10	11
12	7	12
13	15	13
14	11	14
15	16	15
16	6	16
17	9	17
18	9	18
19	5	19
20	8	20
21	9	21
22	6	22
23	3	23
24	9	24
25	8	25

26	4	26
27	3	27
28	2	28
29	3	29
30	2	30
31	5	31
32	4	32
33	2	33
34	8	34
35	6	35
36	6	36
37	6	37
38	1	38
39	4	39
41	3	41
42	3	42
43	4	43
44	3	44
45	3	45
46	2	46
47	2	47
48	4	48
49	2	49
50	4	50
52	4	52
54	2	54
56	2	56
57	1	57
58	3	58
60	4	60
61	1	61
62	1	62
64	1	64
65	1	65
66	1	66
68	2	68
72	1	72
74	2	74
79	2	79
80	1	80
82	1	82
87	1	87
92	1	92
97	2	97
98	1	98
103	3	103
105	1	105
108	1	108
121	1	121
123	1	123
124	1	124
132	3	132
134	1	134
158	1	158

Variable AT, Jobs Between HS and Enlistment

# of Jobs	Frequency	Codes
0	149	0
1	157	1
2	90	2
3	64	3
4	38	4
5	30	5
6	12	6
7	9	7
8	6	8
9	2	9
11	1	11
12	1	12
15	2	15
16	2	16
17	1	17

Variable AU, Months Unemployed HS to Enlistment

# of Months	Frequency	Codes
0	160	0
1	77	1
2	58	2
3	29	3
4	23	4
5	26	5
6	16	6
7	18	7
8	19	8
9	16	9
10	13	10
11	4	11
12	6	12
13	7	13
14	12	14
15	4	15
16	8	16
17	6	17
18	4	18
19	4	19
20	2	20
21	7	21
22	2	22
23	2	23
24	2	24
25	6	25
26	1	26
27	2	27
28	3	28

29	1	29
31	1	31
32	1	32
33	2	33
35	2	35
36	1	36
37	1	37
38	3	38
42	1	42
44	1	44
45	3	45
46	2	46
51	1	51
52	1	52
58	2	58
59	1	59
71	1	71
92	1	92
106	1	106

Variable AV, # Months College Attended

# of Months	Frequency	Codes
0	465	0
1	3	1
2	7	2
3	6	3
4	6	4
5	4	5
6	5	6
7	5	7
8	2	8
9	14	9
10	2	10
11	1	11
12	7	12
13	2	13
14	2	14
15	2	15
16	2	16
18	14	18
20	1	20
21	1	21
22	1	22
24	1	24
26	1	26
27	1	27
28	1	28
29	1	29
36	2	36
42	2	42
44	1	44

45	2	45

Var AW, # Months Unemployed Immediately Prior to Enlistment

# of Months	Frequency	Codes
0	237	0
1	118	1
2	54	2
3	34	3
4	26	4
5	20	5
6	18	6
7	10	7
8	9	8
9	8	9
10	7	10
11	2	11
12	4	12
13	3	13
14	3	14
15	2	15
16	2	16
20	1	20
21	1	21
24	1	24
26	1	26
28	1	28
29	1	29
33	1	33

Variable AX1, Unfavorable Information on SPH #1

Information	Frequency	Codes
Brk & Ent	3	1000
Burglary	3	1001
Larceny	1	1040
Robbery	1	1042
Theft	13	1043
Assault	6	1050
DUI 1	6	1101
DUI >1	1	1102
Drnk & Dis	1	1103
Liq Law Vi	5	1104
Min Alc Po	7	1105
PubIntox 1	2	1106
Marij Arr	1	1120
Marij 1-5	12	1121
Marij >25	1	1123
DrgUse 1-5	1	1141
GED	1	1510

Emplt Prob	1	1533
Tresspass	4	1539
Fired	34	1561
Freq Quit	1	1562
Suspnd Sch	1	1564
Undependbl	1	1578
SusCommSym	1	1600
Leave Acct	2	1804
Reck Drive	1	1805
Mov Veh 1	56	1810
Mov Veh2-5	54	1811
Mov Veh >5	1	1812
Stnd Veh 1	2	1813
StndVeh2-5	2	1814
Disord Con	2	1821
Dstb Peace	1	1822
Harrassmnt	1	1823
Lisc Suspd	1	1825
Loitering	2	1826
Misdemeanr	3	1827
Vandalism	1	1829
Juven Delq	1	1842
Juven Rec	29	1843
Runaway	1	1845
Empl False	2	1924
Time Unact	1	1927
Blank	293	9999

Variable AX2, Unfavorable Information on SPH #2

Information	Frequency	Codes
Brk & Ent	3	1000
Burglary	1	1001
Theft	15	1043
Sus Theft	1	1044
Assault	2	1050
Battery	3	1051
Incorrigbl	1	1070
Alcohol Pr	2	1100
DUI 1	5	1101
Liq Law Vi	3	1104
Min Alc Po	4	1105
PubIntox 1	2	1106
Marij Arr	1	1120
Marij 1-5	2	1121
Marij 6-25	1	1122
Emplt Prob	1	1533
Per Unsat	1	1538
Tresspass	6	1539
Fired	3	1561
Freq Quit	2	1562
Absnt Freq	4	1570
Reck Drive	1	1805

Mov Veh 1	19	1810
Mov Veh2-5	12	1811
Mov Veh >5	1	1812
Disord Con	3	1821
Dstb Peace	3	1822
Il Firearm	1	1824
Lisc Suspd	6	1825
Misdemeanr	1	1827
Crim Misch	1	1840
Juven Rec	4	1843
Mal Mischf	2	1844
Empl False	3	1924

Variable AX3, Unfavorable Information on SPH #3

Information	Frequency	Codes
Larceny	2	1040
Assault	1	1050
Alcohol Pr	1	1100
DUI 1	1	1101
Marij Arr	2	1120
Marij 1-5	5	1121
DrgUse 1-5	1	1141
Non HS Grd	1	1511
Disag Empl	1	1532
Tresspass	1	1539
Absnt Freq	1	1570
Hit & Run	1	1803
Reck Drive	2	1805
Mov Veh 1	7	1810
Mov Veh2-5	2	1811
Stnd Veh 1	1	1813
Disord Con	1	1821
Lisc Suspd	1	1825
Dmg Per Pr	1	1841

Variable AX4, Unfavorable Information on SPH #4

Information	Frequency	Codes
DUI 1	1	1101
Min Alc Po	1	1105
PubIntox 1	1	1106
Bad Checks	1	1230
ContDelMin	1	1410
Leave Acct	1	1804
Mov Veh 1	1	1810
Mov Veh2-5	1	1811
Misdemeanr	1	1827
Juven Rec	1	1843

Mal Mischf	1	1844

Variable AY1, Summary of BI #1

Information	Frequency	Codes
Brk & Ent	1	1000
Burglary	4	1001
Larceny	2	1040
Robbery	1	1042
Theft	9	1043
Sus Theft	3	1044
Assault	3	1050
Rape	1	1052
Lied	3	1071
Violent	1	1090
Alcohol Pr	3	1100
DUI 1	6	1101
DUI >1	1	1102
Liq Law Vi	1	1104
Min Alc Po	2	1105
Marij Arr	2	1120
Marij 1-5	18	1121
Marij 6-25	2	1122
Marij >25	2	1123
Susp Drug	5	1140
DrgUse 1-5	5	1141
Credit	2	1210
Bad Checks	5	1230
Unpd Bills	14	1233
Homo Susp	2	1422
GED	6	1510
Non HS Grd	18	1511
Attitd Pr	1	1530
Emplt Prob	4	1533
Fail Obey	1	1534
Insubordin	1	1535
Not El Reh	1	1537
Perf Unsat	3	1538
Tresspass	3	1539
Tng Dischg	1	1543
Fired	26	1561
Freq Quit	4	1562
Lack Motiv	1	1563
Suspnd Sch	4	1564
Absnt Freq	3	1570
Immat Beha	11	1573
Lack Judge	1	1574
Needs Supv	1	1575
Undependbl	3	1578
Emot Prob	2	1701
Emot Stab?	3	1702
Leave Acct	1	1804
Reck Drive	1	1805

Mov Veh 1	12	1810
Mov Veh2-5	14	1811
Mov Veh >5	2	1812
Runaway	2	1821
Lisc Suspd	1	1825
Misdemeanr	1	1827
Out Warrnt	1	1828
Dmg Per Pr	1	1841
Juven Delq	1	1842
Juven Rec	20	1843
Uncont Juv	1	1846
AddrNotSPH	1	1920
Job NotSPH	3	1923
Empl False	21	1924
Time Unact	3	1927
No Der Inf	282	9999

Variable AY2, Summary of BI #2

Information	Frequency	Codes
Brk & Ent	3	1000
Burglary	1	1001
Forgery	1	1010
Theft	10	1043
Sus Theft	1	1044
Assault	3	1050
Battery	1	1051
Incorrigbl	1	1070
Lied	4	1071
Alcohol Pr	10	1100
DUI 1	5	1101
DUI >1	1	1102
Liq Law Vi	1	1104
Min Alc Po	1	1105
PubIntox 1	1	1106
Marij Arr	2	1120
Marij 1-5	12	1121
Susp Drugs	2	1140
DrgUse 1-5	5	1141
Sale Drugs	4	1150
Drug Waivr	1	1160
Bad Checks	1	1230
No Restit	1	1231
Sued Nonpa	1	1232
Unpd Bills	9	1233
Unpd Taxes	1	1234
Bankruptcy	1	1250
Homo Admit	1	1421
Homo Susp	1	1422
Vio Parole	2	1502
GED	1	1510
Non HS Grd	1	1511
Attitd Pr	3	1530

Emplt Prob	1	1533
Not El Reh	2	1537
Perf Unsat	3	1538
Tresspass	2	1539
Fired	18	1561
Freq Quit	1	1562
Suspd Sch	1	1564
Absnt Freq	10	1570
Braggart	1	1571
Fail Aper	4	1572
Immat Beha	1	1573
Needs Supv	1	1575
Undependbl	3	1578
Emot Prob	1	1701
Emot Stab?	2	1702
Anxiety	1	1710
Depression	2	1723
Reck Drive	1	1805
Mov Veh 1	10	1810
Mov Veh2-5	4	1811
Mov Veh >5	1	1812
StndVeh2-5	1	1814
Lisc Suspd	1	1825
Crim Misch	1	1840
Dmg Per Pr	3	1841
Juven Rec	5	1843
Voyeur	1	1847
Add NotSPH	1	1920
Clmd Diplm	2	1921
Fraud Enli	1	1922
Job NotSPH	1	1923
Empl False	3	1924

Variable AY3, Summary of BI #3

Information	Frequency	Codes
Forgery	1	1010
Larceny	1	1040
R Stol Pro	1	1041
Theft	4	1043
Sus Theft	1	1044
Assault	1	1050
Lied	3	1071
Wild (Beh)	1	1091
Alcohol Pr	6	1100
DUI 1	1	1101
Liq Law Vi	1	1104
Min Alc Po	1	1105
PubIntox 1	2	1106
Marij 1-5	4	1121
Susp Drug	2	1140
DrgUse 1-5	2	1141
Bad Checks	3	1230

No Restit	1	1231
Unpd Bills	3	1233
ContDelMin	1	1410
Un Absense	1	1501
Non HS Grd	1	1511
Not El Reh	3	1537
Perf Unsat	2	1538
Fired	5	1561
Freq Quit	1	1562
Absnt Freq	5	1570
Braggart	1	1571
Immat Beha	2	1573
Tardy	1	1577
Undependbl	2	1578
SusCommSym	1	1600
Emot Prob	1	1701
Reck Drive	1	1805
Mov Veh 1	6	1810
Mov Veh2-5	4	1811
Disord Con	2	1821
Harrassmnt	1	1823
Il Firearm	1	1824
Misdemeanr	1	1827
Out Warrnt	1	1828
Juven Rec	2	1843
Mal Mischf	1	1844
Job NotSPH	1	1923
Empl False	3	1924
Time Unact	2	1927

Variable AY4, Summary of BI #4

Information	Frequency	Codes
Theft	3	1043
Sus Theft	1	1044
Alcohol Pr	1	1100
PubIntox 1	1	1106
Marij 1-5	2	1121
Marij 6-25	1	1122
DrgUse 1-5	3	1141
Not El Reh	1	1537
Fired	2	1561
Suspnd Sch	1	1564
Absnt Freq	2	1570
Immat Beha	3	1573
Reckl Cond	1	1576
Tardy	1	1577
Undependbl	3	1578
Emot Stab?	1	1702
Nerv Cond	1	1712
Leave Acct	1	1804
Reck Drive	1	1805
Mov Veh 1	2	1810

Mov Veh2-5	1	1811
Dstb Peace	1	1822
Empl False	1	1924

Variable BC, Clearance Type

Type of Clearance	Frequency	Codes
No Clrance	3	0
Confident.	1	1
Secret	105	2
Top Secret	141	3
SCI	1	4
Unknown	313	9

Variable BF, Release Code

Release	Frequency	Codes
Pat Miscon	3	1
DrgUse Tra	1	2
DrgUse Pos	12	3
Ser. Offen	2	4
Pers Dis B	7	5
Homos Acts	1	6
Inv Dischg	44	7
HonDisHomo	17	8
Drug Abuse	74	9
Min Dis In	2	10
MisconSerO	14	11
Homos Acts	30	12
Homo State	17	13
HonDisLesb	1	14
Pers Dis	20	15
Con Ct Mar	73	16
HonDisUnsu	39	17
HonDisUnHo	2	18
Vol Dischg	10	19
Invol Rele	162	20
HonDisOffi	33	21

Variable BG1, Military Offenses #1

Information	Frequency	Codes
Larceny	3	1040
Theft	3	1043
Assault	8	1050
Lied	1	1071
Alcohol Pr	1	1100

Drnk & Dis	2	1103
Marij 1-5	44	1121
DrgUse 1-5	32	1141
Sale Drugs	1	1150
Bad Checks	1	1230
Homo Admit	3	1421
Desertion	6	1500
Un Absense	120	1501
Disrespect	6	1530
Derel Duty	2	1531
Fail Obey	25	1534
Disenr Sch	1	1540
Com Threat	1	1820
Disord Con	1	1821
Harrassmnt	1	1823
Fraud Enli	1	1922
No Der Inf	301	9999

Variable BG2, Military Offenses #2

Information	Frequency	Codes
Larceny	1	1040
Theft	3	1043
Assault	8	1050
Alcohol Pr	5	1100
PubIntox 1	1	1106
Marij 1-5	33	1121
Susp Drugs	1	1140
DrgUse 1-5	19	1141
Sale Drugs	2	1150
Att Sodomy	2	1420
Homo Admit	3	1421
Desertion	34	1500
Un Absense	32	1501
Disrespect	4	1530
Derel Duty	2	1531
Fail Obey	24	1534
Emot Stab?	1	1702
Con Weapon	1	1800
Dstb Peace	1	1822
Dmg Per Pr	1	1841

Variable BG3, Military Offenses #3

Information	Frequency	Codes
Larceny	3	1040
Theft	2	1043
Assault	6	1050
Alcohol Pr	5	1100
Drnk & Dis	1	1103

Marij 1-5	13	1121
Susp Drugs	1	1140
DrgUse 1-5	5	1141
Homo Admit	1	1421
Indec Expo	1	1431
Desertion	10	1500
Un Absense	10	1501
Attitd Pr	5	1530
Derel Duty	3	1531
Fail Obey	11	1534
Suicide At	1	1704
Com Threat	1	1820

Variable BG4, Military Offenses #4

Information	Frequency	Codes
Larceny	1	1040
Theft	1	1043
Assault	2	1050
DrgUse 1-5	3	1141
Bad Checks	1	1230
Att Sodomy	1	1420
Attitd Pr	1	1530
Fail Obey	1	1534

Variable BH1, Remarks/Discharge

Remarks	Frequency	Codes
R StolProp	1	1041
Assault	1	1050
Rape	1	1052
Alcohol Pr	2	1100
DUI 1	2	1101
Marij 1-5	1	1121
DrgUse 1-5	5	1141
Drug Waivr	46	1160
Homo Admit	2	1421
Fail Obey	1	1534
Disenr Sch	39	1540
Hard Disch	1	1541
Human Reas	1	1542
Decert PRP	1	1550
Disqul Nuc	7	1551
Emot Stab?	1	1702
Suicide At	2	1704
Disord Con	1	1821
Fraud Enl	11	1922
No Svc Rec	438	9999

Variable BH2, Remarks/Discharge

Remarks	Frequency	Codes
Burglary	1	1001
Larceny	1	1040
Lied	1	1071
Alcohol Pr	2	1100
Marij 1-5	1	1121
DrgUse 1-5	8	1141
Sale Drugs	1	1150
Homo Admit	3	1421
Un Absense	1	1501
Disenr Sch	7	1540
Decert PRP	6	1550
Disqul Nuc	5	1551
Lack Motiv	3	1563
Suicide At	1	1704
Disord Con	1	1821
Fraud Enli	2	1922

Variable BH3, Remarks/Discharge

Remarks	Frequency	Codes
Sale Drugs	1	1050
PubIntox 1	1	1106
DrgUse 1-5	2	1141
Un Absense	1	1501
Disenr Sch	1	1540
Decert PRP	2	1550
Disqul Nuc	1	1551
Harrassmnt	1	1823

Variable BH4, Remarks/Discharge

Remarks	Frequency	Codes
Homo Admit	1	1421

Variable BL, Status of 5520/20

Status	Frequency	Codes
No Rec BI	82	90
Rec Not Fd	70	98
Blank	412	99

Variable BM, Discharge Case Codes

Case Codes	Frequency	Codes
BI-Clear	324	101
SBI-Clear	119	102
IBI-Clear	35	103
BI-Suitabi	74	113
SBI-Suitab	12	123

Variable BO, Interservice Seperation Code

Seperation Codes	Frequency	Codes
Unknown	10	0
ETS	195	1
Char/Beh D	65	60
Discred In	50	65
Drugs	86	67
Court Mart	73	73
Homosexual	70	76
Serious Of	15	84

Variable BP, Type of Discharge

Type	Frequency	Codes
Honorable	328	1
General	10	2
Oth Th Hon	226	3

Variable BQ, Character of Service

Character	Frequency	Codes
Good	205	0
Homosexual	70	1
Misconduct	65	2
Drug/Alcoh	86	3
Court Mart	73	4
Char/Beh D	65	5

APPENDIX B. FOUR-DIGIT DEROGATORY INFORMATION CODES

The codes contained in this appendix are those used in the data base.

CODE	Behavior
1000	Breaking and entering
1001	Burglary
1010	Forgery
1040	Larceny
1041	Received stolen property
1042	Robbery
1043	Theft, grand or petty
1044	Theft suspected
1050	Assault
1051	Battery
1052	Rape
1060	Dishonest
1070	Incorrigible
1071	Lied
1090	Violent
1091	Wild (behavior)
1100	Alcohol problems
1101	Driving under the influence, 1 time
1102	Driving under the influence, more than 1 time
1103	Drunk and disorderly
1104	Liquor law violation
1105	Minor in possession of alcohol
1106	Public Intoxication, 1 time
1107	Public Intoxication, more than 1 time
1120	Marijuana arrest
1121	Marijuana use 1 to 5 times
1122	Marijuana use 6 to 25 times
1123	Marijuana use more than 25 times
1140	Drug use suspected
1141	Drug use (cocaine, heroin, etc.) 1 to 5 times
1142	Drug use (cocaine, heroin, etc.) 6 to 25 times
1143	Drug use (cocaine, heroin, etc.) more than 25 times
1150	Sale of drugs
1160	Drug Waiver
1210	Credit
1230	Bad checks
1231	Restitution not made
1232	Sued for non-payment of account
1233	Unpaid bills or accounts
1234	Unpaid taxes
1235	Vehicle repossessed
1250	Bankrupcy
1400	Do anything for money
1410	Contributing to the delinquency of a minor
1420	Attempted Sodomy
1421	Admitted homosexual

1422 Suspected homosexual
1423 Juvenile homosexual behavior
1430 Dressed as a female
1431 Indecent Exposure
1500 Desertion
1501 Unauthorized absense
1502 Violation of parole
1510 General Education Degree (GED)
1511 Did not graduate from high school
1530 Attitude Problem
1530 Disrespect (same code as above)
1531 Dereliction of duty
1532 Disagreements with employer
1533 Employment problem
1534 Failure to obey a lawful order
1535 Insubordinate
1536 Misconduct
1537 Not eligible for rehire
1538 Performance unsatisfactory
1539 Tresspass
1540 Disenrolled from school
1541 Hardship discharge required
1542 Humanitarian reassignment
1543 Training discharge
1550 Decertified from Personnel Reliability Program (PRP)
1551 Disqualified from the nuclear field
1560 Expelled
1561 Fired
1562 Frequently quit or left job
1563 Lack of motivation
1564 Suspended from school
1570 Absent frequently
1571 Braggart
1572 Failure to appear
1573 Immature behavior
1574 Lacked judgement
1575 Needs supervision
1576 Reckless conduct
1577 Tardy (school or job)
1578 Undependable
1600 Suspected communist sympathies
1700 Easily upset
1701 Emotional problems
1702 Mental or emotional stability questioned
1703 Psychological problem
1704 Suicide attempt
1710 Anxiety
1711 Introverted
1712 Nervous condition or unable to cope with stress
1723 Depression
1800 Concealed weapon
1801 Felony
1802 Fugitive
1803 Hit and run
1804 Leaving the scene of an accident
1805 Reckless driving

1810 Vehicle violation, moving, 1
1811 Vehicle violation, moving, 2 to 5
1812 Vehicle violation, moving, 6 or more
1813 Vehicle violation, standing, 1
1814 Vehicle violation, standing, 2 to 5
1815 Vehicle violation, standing, 6 or more
1820 Communicate a threat
1821 Disorderly conduct
1822 Disturbing the peace
1823 Harrassment
1824 Illegal use of firearms
1825 Liscense suspended
1826 Loitering
1827 Misdemeanor
1828 Outstanding arrest warrant
1829 Vandalism
1840 Criminal mischief
1841 Damage to personal property
1842 Juvenile delinquent
1843 Juvenile record
1844 Malicious mischief
1845 Runaway
1846 Uncontrolled juvenile
1847 Voyeur or Peeping Tom
1900 Used alias
1920 Address not listed on Statement of Personal History (SPH)
1921 Claimed to be a diploma graduate of high school
1922 Fraudulent reenlistment
1923 Jobs not listed on SPH
1924 Length of employment falsified
1925 Prior service concealed
1926 Social Security Number (SSN) falsified or altered
1927 Time unaccounted for

APPENDIX C. BOOTSTRAP RESIDUAL PROGRAM

This appendix contains the program which produces an empirical estimate of the upper-quantile distribution of the maximum residuals produced by independent, cross-tabulated variables.

```
      PROGRAM BSRESD
*************************************************************************
*     THIS PROGRAM CONDUCTS A MONTE-CARLO SIMULATION USING THE MULTI-   *
*     NOMIAL RANDOM NUMBER GENERATOR FROM THE IMSL LIBRARY.  THE        *
*     PURPOSE OF THE SIMULATION IS TO GENERATE A NUMBER OF REPLICATIONS*
*     OF POSSIBLE COUNTS OCCURING IN A CONTINGENCY TABLE WHEN THE       *
*     NULL HYPOTHESIS OF INDEPENDENCE, OR EQUAL PROBABILITY APPLIES.    *
*     SPECIFIC ROW AND COLUMN COUNTS FROM CONTINGENCY TABLES, AND A     *
*     HEADING FOR IDENTIFICATION ARE READ FROM AN INPUT FILE.  THE      *
*     MULTINOMIAL RANDOM NUMBER GENERATOR PRODUCES SIMULATED COUNTS     *
*     BASED UPON THE CELL PROBABILITIES PRODUCED FROM THE MARGINAL      *
*     COUNTS.  EACH REPLICATION OF THE SIMULATION IS TREATED AS A       *
*     NEW CONTINGENCY TABLES AND THE STANDARDIZED RESIDUALS ARE COM-    *
*     PUTED AS IN AN ACTUAL CONTINGENCY TABLE.  THE LARGEST ABSOLUTE    *
*     RESIDUAL FOR EACH REPLICATION IS SAVED AND STATISTICAL            *
*     INFORMATION ABOUT THE LARGEST ABSOLUTE RESIDUAL PRODUCED BY       *
*     INDEPENDENT VARIABLES IS COMPUTED.                                *
*                                                                       *
*************************************************************************
      CHARACTER*79 TBLNAM
      REAL MXRESD(200),ORDMRE(200),RESID(200,70),MNRSD,P(70),E(70)
      INTEGER ROWNUM,COLNUM,ROWCNT(20),COLCNT(20),TOTAL,RSLT(200,70),
     X   CELLS

*ESTABLISHES THE INPUT/OUTPUT FILES
      CALL EXCMS('FILEDEF 01 DISK BSRESID DATA A')
      CALL EXCMS('FILEDEF 02 DISK BSRESID OUTPUT A')

      NR = 200
      M = 1

*SETS THE SEED FOR THE RANDOM NUMBER GENERATOR
      ISEED = 54821
      CALL RNSET(ISEED)

152   FORMAT(1X,'STATISTICS OF THE MAX RESIDUALS DERIVED FROM A ',
     X   'BOOTSTRAP'/1X,'SIMULATION OF CROSSTABULATED INDEPENDENT ',
     X   'CATEGORICAL VARIABLES.'//)
      WRITE(2,152)

99    READ(1,100,END=999)TBLNAM
100   FORMAT(A)
      READ(1,110)ROWNUM,COLNUM
110   FORMAT(2I5)
      READ(1,120)(ROWCNT(I),I=1,ROWNUM)
```

```
      READ(1,120)(COLCNT(I),I=1,COLNUM)
120   FORMAT(20I4)

      TOTAL = 0
      DO 10 I=1,COLNUM
         TOTAL = TOTAL + COLCNT(I)
10    CONTINUE

*ESTABLISHES THE EXPECTED CELL SIZE AND THE PROBABILITY OF A COUNT
*FALLING IN EACH CELL

      PTOT = 0.
      K = 1
      DO 20 I=1,ROWNUM
         DO 20 J=1,COLNUM
              E(K) = (1.0 * ROWCNT(I) * COLCNT(J)) / (1.0 * TOTAL)
              P(K) = E(K) / TOTAL
              PTOT = PTOT + P(K)
              K = K + 1
20    CONTINUE

      CELLS = ROWNUM * COLNUM

*ERROR TRAP IN CASE ROUNDING ERROR CAUSES PROBLEMS

      IF (ABS(1.0 - PTOT) .GT.002) THEN
         WRITE(2,*) ' LARGE DISCREPANCY'
         GO TO 99
      END IF

      IF (PTOT .NE. 1.0) THEN
         P(CELLS) = P(CELLS) + (1.0 - PTOT)
      END IF

*CALLS THE MULTINOMIAL RANDOM NUMBER GENERATOR

      CALL RNMTN(NR,TOTAL,CELLS,P,RSLT,200)

*COMPUTES THE RESIDUALS FOR EACH REPLICATION AND EACH CELL

      DO 30 I=1,NR
         MXRESD(I) = 0.
         DO 30 J=1,CELLS
              RESID(I,J) = ( RSLT(I,J) - E(J) ) / ( SQRT( E(J) ) )
              IF( ABS( RESID(I,J) ) .GT. MXRESD(I) )
     X             MXRESD(I) = ABS( RESID(I,J) )
30    CONTINUE

*USES A BUBBLE SORT ROUTINE TO ARRANGE FROM SMALLEST TO LARGEST

      CALL BUBBLE(MXRESD,NR,ORDMRE)

*FORMATTED OUTPUT FOR OUTPUT FILE

      WRITE(2,101)TBLNAM
101   FORMAT (1X,A)
```

```
      WRITE(2,151)M,ROWNUM,COLNUM
151   FORMAT(1X,'RECORD NUMBER = ',I3/1X,'A ',I2,' BY ',I2,' TABLE '/
     X   1X,'WITH THE FOLLOWING EXPECTED CELL SIZES:'/)
      K = 0
      DO 35 I=1,ROWNUM
         WRITE(2,153)(E(J),J=K+1,K+COLNUM)
         K = K + COLNUM
35    CONTINUE
153   FORMAT(15(1X,F5.1))

*WRITES OUT THE STATISTICS

      MNSRD = 0.
      DO 50 I=1,NR
         MNRSD = ORDMRE(I) + MNRSD
50    CONTINUE

      MNRSD = MNRSD / NR

      DO 60 I=1,NR
         VARRSD = (ORDMRE(I) - MNRSD)**2
60    CONTINUE

      VARRSD = VARRSD / NR
      SDRSD = SQRT (VARRSD)

      WRITE(2,160)

160   FORMAT(/T4,'MEAN',T13,'VARIANCE',T24,'STD DEV',T35,'.95 QUANT',
     X   T48,'.99 QUANT',T61,'LARGEST VALUE')

      WRITE(2,161)MNRSD,VARRSD,SDRSD,ORDMRE(INT(.95*NR)),ORDMRE(INT(.99*
     X   NR)),ORDMRE(NR)
161   FORMAT(T4,F6.4,T13,F6.4,T24,F6.4,T35,F6.4,T48,F6.4,T61,F6.4//)

      GO TO 99
999   STOP
      END

      SUBROUTINE BUBBLE(ARG,N,RSLT)

*USES THE BUBBLE SORT ALGORITHM TO ARRANGE A VECTOR OF VALUES, ARG

      REAL ARG(N),TEMP,RSLT(N)

      DO 10 I=1,N-1
         DO 10 J =N,I+1,-1
              IF (ARG(J).LT.ARG(J-1)) THEN
                   TEMP = ARG(J)
                   ARG(J) = ARG(J-1)
                   ARG(J-1) = TEMP
              END IF
10    CONTINUE

      DO 20 I=1,N
```

```
         RSLT(I) = ARG(I)
20    CONTINUE

      END
```

APPENDIX D. INPUT FILE FOR THE BOOTSTRAP SIMULATION PROGRAM

This appendix contains the input for each contingency table. Each table is represented by four lines. Line 1 is the title; line 2 is the number of rows and then the number of columns; line 3 is the row marginal counts; line 4 is the column marginal counts.

```
RACE VS. CHARACTER OF SERVICE
    4    6
 397  50  13 104
 205  70  65  86  73  65
SUBJECTS WITH AN INCIDENT OF MAJOR CRIME VS. CHARACTER OF SERVICE
    2    6
 444 120
 205  70  65  86  73  65
SUBJECTS WITH ADJUSTMENT INCIDENTS DISCLOSED IN INV VS. CHAR OF SVC
    2    6
 336 228
 205  70  65  86  73  65
MARITAL STATUS VS. CHARACTER OF SERVICE
    3    6
 517  33  14
 205  70  65  86  73  65
SUBJECT SEX VS. CHARACTER OF SERVICE
    2    6
 479  85
 205  70  65  86  73  65
HIGH SCHOOL DIPLOMA BEARER VS. CHARACTER OF SERVICE
    2    6
 454 110
 205  70  65  86  73  65
CATEGORY OF MOST SERIOUS DRUG USE VS. CHARACTER OF SERVICE
    3    6
 418 105  41
 205  70  65  86  73  65
DRUG/ALCOHOL ABUSE DISCLOSED IN INVESTIGATION VS. CHARACTER OF SERVICE
    2    6
 389 175
 205  70  65  86  73  65
AMOUNT OF DRUG USE DISCLOSED IN INVESTIGATION VS. CHARACTER OF SERVICE
    4    6
 418  64  35  47
 205  70  65  86  73  65
DRUG USE DISCLOSED IN INVESTIGATION VS. CHARACTER OF SERVICE
    2    6
 418 146
 205  70  65  86  73  65
AGE AT ENLISTMENT VS. CHARACTER OF SERVICE
    4    6
  85 382  78  19
```

```
 205  70  65  86  73  65
AGE AT ENLISTMENT VS. CHARACTER OF SERVICE
    3    6
  85 382  97
 205  70  65  86  73  65
DEROGATORY INFO DISCLOSED IN INVESTIGATION VS. CHARACTER OF SERVICE
    2    2
 151 413
 205 359
DEROGATORY INFORMATION DISCLOSED IN INVESTIGATION VS. CHAR. OF SERVICE
    2    6
 151 413
 205  70  65  86  73  65
CHAR OF SVC VS. AMOUNT OF DEROG. INFORMATION DISCLOSED IN INVESTIGATION
    6   11
 205  70  65  86  73  65
 151  65  68  36  25  28  32  26  25  29  79
CHAR OF SVC VS. WEIGHTED RECOMMENDATION SCORE
    6    4
 205  70  65  86  73  65
 105 103 284  72
CHAR OF SVC VS. RECOMMENDATION SCORE (ADJUSTED FOR NEGATIVE RECS.)
    6    4
 205  70  65  86  73  65
  82 110 296  76
CHAR OF SVC VS. NUMBER OF RECOMMENDATIONS DISCLOSED IN INVESTIGATION
    6    5
 205  70  65  86  73  65
  62  84 314  67  36
CHAR OF SVC VS. NUMBER OF RECOMMENDATIONS DISCLOSED IN INVESTIGATION
    6    4
 205  70  65  86  73  65
  62  84 314 104
CHAR OF SVC VS. NUMBER OF NOT RECOMMENDED DISCLOSED IN INVESTIGATION
    6    2
 205  70  65  86  73  65
 489  75
ADMITTED DRUG USE (SPH) VS. DRUG USE DISCLOSED BY LOCAL AGENCY CHECK
    2    2
 504  60
 510  54
INFO OBTAINED FROM INTERVIEW (YES OR NO) VS. REASON FOR INTERVIEW
    2    4
  93 195
  79  15 118  76
AMOUNT OF INFO OBTAINED FROM SUBJECT INTERVIEW VS. REASON FOR INTERVIEW
    5    4
  93  72  66  33  24
 79  15 118  76
```

APPENDIX E. OUTPUT FROM THE BOOTSTRAP SIMULATION OF THE LARGEST RESIDUAL

This appendix contains the output from the FORTRAN program which extracted the maximum residual from each replication of independent cross-tabulated variables.

```
STATISTICS OF THE MAX RESIDUALS DERIVED FROM A BOOTSTRAP
SIMULATION OF CROSSTABULATED INDEPENDENT CATEGORICAL VARIABLES.


RACE VS. CHARACTER OF SERVICE
RECORD NUMBER =    1
A  4 BY  6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

144.3  49.3  45.8  60.5  51.4  45.8
 18.2   6.2   5.8   7.6   6.5   5.8
  4.7   1.6   1.5   2.0   1.7   1.5
 37.8  12.9  12.0  15.9  13.5  12.0

  MEAN      VARIANCE    STD DEV     .95 QUANT     .99 QUANT     LARGEST VALUE
  2.1611    0.0126      0.1120      2.8873        3.6441        3.7455


SUBJECTS WITH AN INCIDENT OF MAJOR CRIME VS. CHARACTER OF SERVICE
RECORD NUMBER =    1
A  2 BY  6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

161.4  55.1  51.2  67.7  57.5  51.2
 43.6  14.9  13.8  18.3  15.5  13.8

  MEAN      VARIANCE    STD DEV     .95 QUANT     .99 QUANT     LARGEST VALUE
  1.9063    0.0236      0.1536      2.8432        3.3313        4.0793


SUBJECTS WITH ADJUSTMENT INCIDENTS DISCLOSED IN INV VS. CHAR OF SVC
RECORD NUMBER =    1
A  2 BY  6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

122.1  41.7  38.7  51.2  43.5  38.7
 82.9  28.3  26.3  34.8  29.5  26.3

  MEAN      VARIANCE    STD DEV     .95 QUANT     .99 QUANT     LARGEST VALUE
  1.8898    0.0115      0.1070      2.6711        3.1063        3.4036


MARITAL STATUS VS. CHARACTER OF SERVICE
RECORD NUMBER =    1
```

A 3 BY 6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

187.9	64.2	59.6	78.8	66.9	59.6
12.0	4.1	3.8	5.0	4.3	3.8
5.1	1.7	1.6	2.1	1.8	1.6

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
2.0199	0.0332	0.1823	3.1063	3.9979	4.5982

SUBJECT SEX VS. CHARACTER OF SERVICE
RECORD NUMBER = 1
A 2 BY 6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

174.1	59.5	55.2	73.0	62.0	55.2
30.9	10.5	9.8	13.0	11.0	9.8

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
1.8686	0.0267	0.1633	2.7139	3.0773	4.1773

HIGH SCHOOL DIPLOMA BEARER VS. CHARACTER OF SERVICE
RECORD NUMBER = 1
A 2 BY 6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

165.0	56.3	52.3	69.2	58.8	52.3
40.0	13.7	12.7	16.8	14.2	12.7

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
1.8848	0.0257	0.1604	2.6857	3.1658	4.1537

CATEGORY OF MOST SERIOUS DRUG USE VS. CHARACTER OF SERVICE
RECORD NUMBER = 1
A 3 BY 6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

151.9	51.9	48.2	63.7	54.1	48.2
38.2	13.0	12.1	16.0	13.6	12.1
14.9	5.1	4.7	6.3	5.3	4.7

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
2.1042	0.0170	0.1305	2.9055	3.7080	3.9504

DRUG/ALCOHOL ABUSE DISCLOSED IN INVESTIGATION VS. CHARACTER OF SERVICE
RECORD NUMBER = 1
A 2 BY 6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

141.4	48.3	44.8	59.3	50.3	44.8
63.6	21.7	20.2	26.7	22.7	20.2

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
1.8342	0.0113	0.1061	2.5641	2.9649	3.3349

AMOUNT OF DRUG USE DISCLOSED IN INVESTIGATION VS. CHARACTER OF SERVICE
RECORD NUMBER = 1
A 4 BY 6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

151.9	51.9	48.2	63.7	54.1	48.2
23.3	7.9	7.4	9.8	8.3	7.4
12.7	4.3	4.0	5.3	4.5	4.0
17.1	5.8	5.4	7.2	6.1	5.4

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
2.2402	0.0370	0.1925	3.1621	3.7233	4.9623

DRUG USE DISCLOSED IN INVESTIGATION VS. CHARACTER OF SERVICE
RECORD NUMBER = 1
A 2 BY 6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

151.9	51.9	48.2	63.7	54.1	48.2
53.1	18.1	16.8	22.3	18.9	16.8

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
1.8495	0.0471	0.2170	2.5684	3.1447	4.9181

AGE AT ENLISTMENT VS. CHARACTER OF SERVICE
RECORD NUMBER = 1
A 4 BY 6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

30.9	10.5	9.8	13.0	11.0	9.8
138.8	47.4	44.0	58.2	49.4	44.0
28.4	9.7	9.0	11.9	10.1	9.0
6.9	2.4	2.2	2.9	2.5	2.2

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
2.1758	0.0112	0.1059	3.0228	3.5332	3.6740

AGE AT ENLISTMENT VS. CHARACTER OF SERVICE
RECORD NUMBER = 1
A 3 BY 6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

30.9	10.5	9.8	13.0	11.0	9.8
138.8	47.4	44.0	58.2	49.4	44.0
35.3	12.0	11.2	14.8	12.6	11.2

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
2.0881	0.0164	0.1281	2.9407	3.6218	3.8992

DEROGATORY INFO DISCLOSED IN INVESTIGATION VS. CHARACTER OF SERVICE
RECORD NUMBER = 1
A 2 BY 2 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

54.9	96.1
150.1	262.9

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
1.1540	0.0117	0.1082	1.8866	2.3578	2.6841

DEROGATORY INFORMATION DISCLOSED IN INVESTIGATION VS. CHAR. OF SERVICE
RECORD NUMBER = 1
A 2 BY 6 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

54.9	18.7	17.4	23.0	19.5	17.4
150.1	51.3	47.6	63.0	53.5	47.6

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
1.9345	0.0163	0.1276	2.9014	3.4961	3.7390

CHAR OF SVC VS. AMOUNT OF DEROG. INFORMATION DISCLOSED IN INVESTIGATION
RECORD NUMBER = 1
A 6 BY 11 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

54.9	23.6	24.7	13.1	9.1	10.2	11.6	9.5	9.1	10.5	28.7
18.7	8.1	8.4	4.5	3.1	3.5	4.0	3.2	3.1	3.6	9.8
17.4	7.5	7.8	4.1	2.9	3.2	3.7	3.0	2.9	3.3	9.1
23.0	9.9	10.4	5.5	3.8	4.3	4.9	4.0	3.8	4.4	12.0
19.5	8.4	8.8	4.7	3.2	3.6	4.1	3.4	3.2	3.8	10.2
17.4	7.5	7.8	4.1	2.9	3.2	3.7	3.0	2.9	3.3	9.1

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
2.7242	0.0226	0.1502	3.7704	4.6236	4.8490

CHAR OF SVC VS. WEIGHTED RECOMMENDATION SCORE
RECORD NUMBER = 1
A 6 BY 4 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

38.2	37.4	103.2	26.2
13.0	12.8	35.2	8.9
12.1	11.9	32.7	8.3
16.0	15.7	43.3	11.0
13.6	13.3	36.8	9.3
12.1	11.9	32.7	8.3

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
2.2328	0.0233	0.1526	3.0383	3.7152	4.3912

CHAR OF SVC VS. RECOMMENDATION SCORE (ADJUSTED FOR NEGATIVE RECS.)
RECORD NUMBER = 1
A 6 BY 4 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

29.8	40.0	107.6	27.6
10.2	13.7	36.7	9.4
9.5	12.7	34.1	8.8
12.5	16.8	45.1	11.6
10.6	14.2	38.3	9.8
9.5	12.7	34.1	8.8

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
2.1837	0.0125	0.1119	2.9216	3.4407	3.7663

CHAR OF SVC VS. NUMBER OF RECOMMENDATIONS DISCLOSED IN INVESTIGATION
RECORD NUMBER = 1
A 6 BY 5 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

22.6	30.6	114.3	24.4	13.1
7.7	10.4	39.0	8.3	4.5
7.2	9.7	36.3	7.7	4.2
9.5	12.8	48.0	10.2	5.5
8.0	10.9	40.7	8.7	4.7
7.2	9.7	36.3	7.7	4.2

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
2.3241	0.0291	0.1706	3.1986	4.0515	4.7361

CHAR OF SVC VS. NUMBER OF RECOMMENDATIONS DISCLOSED IN INVESTIGATION
RECORD NUMBER = 1
A 6 BY 4 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

22.5	30.5	114.1	37.8
7.7	10.4	39.0	12.9
7.1	9.7	36.2	12.0
9.5	12.8	47.9	15.9
8.0	10.9	40.6	13.5
7.1	9.7	36.2	12.0

MEAN	VARIANCE	STD DEV	.95 QUANT	.99 QUANT	LARGEST VALUE
2.1945	0.0172	0.1311	3.1814	3.4176	4.0479

CHAR OF SVC VS. NUMBER OF NOT RECOMMENDED DISCLOSED IN INVESTIGATION
RECORD NUMBER = 1
A 6 BY 2 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

177.7	27.3
60.7	9.3

```
 56.4    8.6
 74.6   11.4
 63.3    9.7
 56.4    8.6

  MEAN      VARIANCE    STD DEV     .95 QUANT     .99 QUANT     LARGEST VALUE
  1.9180    0.0161      0.1271      2.8423        3.3049        3.7152


ADMITTED DRUG USE (SPH) VS. DRUG USE DISCLOSED BY LOCAL AGENCY CHECK
RECORD NUMBER =    1
A  2 BY  2 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

455.7   48.3
 54.3    5.7

  MEAN      VARIANCE    STD DEV     .95 QUANT     .99 QUANT     LARGEST VALUE
  1.3037    0.0327      0.1809      2.1961        3.0271        3.8615


INFO OBTAINED FROM INTERVIEW (YES OR NO) VS. REASON FOR INTERVIEW
RECORD NUMBER =    1
A  2 BY  4 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

 25.5    4.8   38.1   24.5
 53.5   10.2   79.9   51.5

  MEAN      VARIANCE    STD DEV     .95 QUANT     .99 QUANT     LARGEST VALUE
  1.6533    0.0103      0.1015      2.4613        2.7335        3.0888


AMOUNT OF INFO OBTAINED FROM SUBJECT INTERVIEW VS. REASON FOR INTERVIEW
RECORD NUMBER =    1
A  5 BY  4 TABLE
WITH THE FOLLOWING EXPECTED CELL SIZES:

 25.5    4.8   38.1   24.5
 19.8    3.8   29.5   19.0
 18.1    3.4   27.0   17.4
  9.1    1.7   13.5    8.7
  6.6    1.3    9.8    6.3

  MEAN      VARIANCE    STD DEV     .95 QUANT     .99 QUANT     LARGEST VALUE
  2.1487    0.0349      0.1868      3.2275        3.9712        4.7911
```

LIST OF REFERENCES

1. *DOD Personal Security Regulation,* DOD 5200.2-R, Washington, DC, January, 1987.

2. Collot, Patricia L. and Laurence, Janice H., Comparison in Evaluation of Interview Oriented and Special Background Investigations, final report, 87-03, HumRRO International, Alexandria, VA, November 1987.

3. Conover, W.J., *Practical Nonparametric Statistics,* John Wiley & Sons, Inc., New York, NY, 1980.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	PERSEREC Naval Postgraduate School, Code 65 Monterey, CA 93943-5000	17
4.	Professor Peter A.W. Lewis Department of Operations Research, Code 55 Naval Postgraduate School Monterey, CA 93943-5000	5
5.	Dr. Peter Welch IBM Research Center P.O. Box 2181 Yorktown Heights, NY 10599	2
6.	CPT Edward R. Koucheravy Mathematics Department U.S. Military Academy West Point, NY 10996	2